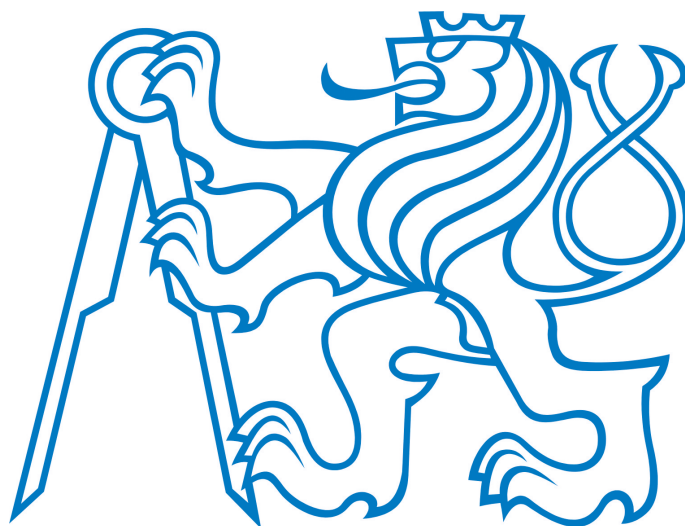


České vysoké učení technické v Praze
Fakulta elektrotechnická



Objektivní hodnocení kvality
stereoskopického obrazu

Diplomová práce

Praha, 2013

Autor: Bc. Jan Dlouhý

Vedoucí práce: Ing. Karel Fliegel Ph.D.

Prohlášení

Prohlašuji, že jsem diplomovou práci Objektivní hodnocení kvality stereoskopického obrazu vypracoval samostatně a použil k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k diplomové práci.

Nemám námitky proti půjčování, zveřejnění a dalšímu využití práce, pokud s tím bude souhlasit katedra radioelektroniky.

V Praze dne _____

_____ podpis

Abstrakt

Náplní této práce je vypracování přehledu dostupných metod na hodnocení kvality stereoskopického obrazu. Byla provedena implementace těchto metod: Depth Perception and Motion Cue based 3D Video Quality Assessment, Stereoscopic video quality perception model a Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction. Účinnost implementací byla ověřena na dvou dostupných databázích. Jedná se o databázi NAMA3DS1 - COSPAD1 vytvořenou institutem IRCCyN na universitě Nantes a databázi vytvořenou skupinou MMSPG na Švýcarském federálním technologickém institutu v Lausanne. Výsledky testů jsou poté zhodnoceny.

Abstract

The aim of this work is to develop an overview of available quality assessment methods of stereoscopic images and video. Following methods were implemented: Depth Perception and Motion Cue based 3D Video Quality Assessment, Stereoscopic video quality perception model and Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction. The effectiveness of the implementations was verified on two available databases: NAMA3DS1 - COSPAD1 created by IRCCyN Insitute at the University of Nantes and a database created by MMSPG group at the Swiss Federal Institute of Technology in Lausanne. The test results are then evaluated.

Obsah

Seznam obrázků	v
Seznam tabulek	vi
1 Úvod	1
1.1 Artefakty a zkreslení typická pro stereoskopické video	2
1.2 Přehled používaných formátů a standardů	3
2 Metody s referencí	6
2.1 Metody založené na SSIM - Structural similarity index metric/index strukturní podobnosti	6
2.1.1 Using disparity for quality assesment of stereoscopic images - 2008 . .	7
2.1.2 Quality evaluation of strereoscopic videos using depth map segmentation - 2011	8
2.1.3 Depth Perception and Motion Cue based 3D Video Quality Assesment - 2012	9
2.1.4 Assessment of Stereoscopic Crosstalk Perception	11
2.2 Metody založené na MSE - Mean square error	12
2.2.1 3D-DCT based perceptual quality assessment of stereo video 2011 . .	12
3 Metody bez reference	14
3.1 Stereoscopic video quality perception model - SV-QPM 2011	14
3.1.1 Disparity	14
3.1.2 Časová změna pohybového vektoru	15
3.1.3 Variace hodnot disparity v rámci snímku (Intra-frame disparity variation)	16
3.1.4 Mezišnímková variace hodnot disparity (Inter-frame disparity variation)	16
3.1.5 Rozložení disparity v hraničních oblastech snímku (Disparity distribution of frame boundary area)	17
3.1.6 Stereoscopic video quality perception model (SV-QPM)	17

3.2	Algorithmic assesment of 3D quality of experience for images and videos - 2011	17
3.2.1	Disparity map a její statistické parametry	18
3.2.2	Spatial activity a její statistické parametry	18
3.2.3	Diferenční disparity map a její statistické parametry	19
3.3	A no-reference quality measure for DIBR-based 3D videos - 2011	20
3.3.1	Odhad ideální depth map	20
3.3.2	Parametry charakterizující zkreslení	21
3.3.3	Hodnocení kvality - NR-3VQM	22
3.4	Spatio-temporal segmentation based continuous no-reference stereoscopic vi- deo quality prediction 2010	22
3.4.1	Segmentace na edge/no-edge bloky	23
3.4.2	Artifacts measure - měření artefaktů	23
3.4.3	Disparity measure	25
3.4.4	Temporal features	26
3.4.5	Hodnocení kvality - STS	26
4	Databáze	28
4.1	Datáze EPFL	28
4.2	Databáze NAMA3DS1-COSPAD1	29
5	Implementace a Výsledky	30
5.1	Using disparity for quality assesment of stereoscopic images - 2008	30
5.2	Stereoscopic video quality perception model - 2011	32
5.3	Spatio-temporal segmentation based continuous no-reference stereoscopic vi- deo quality prediction 2010	34
6	Závěr	38
	Literatura	43
	A Obsah přiloženého CD	I
	B Přehled použitých skriptů v jazyce MATLAB	II

Seznam obrázků

1.1	Obraz a jeho depth map [15]	4
2.1	Blokové schéma metody [20]	7
2.2	Blokové schéma metody [26]	9
5.1	Znázornění korelace mezi DMOS a dílčími parametry, databáze EPFL	31
5.2	Znázornění korelace mezi DMOS a dílčími parametry, databáze NAMA3DS1	31
5.3	DMOS vs DMOS_p, (a) - EPFL, (b) NAMA3DS1	32
5.4	MOS vs SVQPAM, databáze EPFL	33
5.5	MOS vs SVQPAM, databáze NAMA3DS1	33
5.6	MOS vs MOS_p, (a) - EPFL, (b) NAMA3DS1	34
5.7	Znázornění korelace mezi MOS a dílčími parametry metody, edge area, data- báze NAMA3DS1	35
5.8	Znázornění korelace mezi MOS a dílčími parametry metody, edge area, data- báze EPFL	35
5.9	Znázornění korelace mezi DMOS a dílčími parametry metody, databáze NAMA3DS1	36
5.10	MOS vs MOS_p, (a) - EPFL, (b) NAMA3DS1	37

Seznam tabulek

3.1	PCC - NR-3VQM	22
5.1	Porovnání odhadnutých výsledků se subjektivními testy	32
5.2	Porovnání odhadnutých výsledků se subjektivními testy - SVQPM	34
5.3	Regresní koeficienty	36
5.4	Výsledky metody	37

Kapitola 1

Úvod

K velkému rozvoji trojrozměrné reprezentace obrazu dochází především ve snaze o lepší zážitek a více realistickou reprezentaci. Nejrozšířenější metodou, která navozuje dojem trojrozměrného prostoru je stereoskopické video. Tato metoda je založena na principu stereopse. Člověk má oči vzdálené od sebe přibližně 6 cm, každým okem tedy vnímá scénu z jiného úhlu [1]. Rozdílu mezi těmito obrazy se říká binocular disparity. Spojením těchto obrazů v CNS vzniká vjem hloubky [2][3]. Pokud tedy dokážeme při zobrazení videa každému oku promítnout jiný obraz můžeme navodit vjem trojrozměrného zobrazení.

S rozšířením technologie stereoskopického videa dochází k vývoji nových metod snímání, komprese, kódování, přenosu, zobrazení atd. V každém z těchto procesů dochází k degradaci videa. Hodnocení kvality je tedy při návrhu těchto procesů zásadní. Nejspolehlivějším ohodnocením jsou subjektivní testy. Ty jsou ale drahé, zdlouhavé a nehodí se pro online aplikace [4][5]. Vývoj metod pro objektivní hodnocení kvality je tedy nezbytný.

Metody se podle míry reference dělí na metody s referencí, s redukovanou referencí a bez reference [6][7].

Metody s referencí používají pro hodnocení kvality porovnání s referenčním (nezkresleným) videem. Při jejich vývoji se většinou vycházelo z dosud známých metod pro hodnocení kvality videa.

Metody s redukovanou referencí využívají podobných principů jako metody s referencí. Pouze dochází k redukci informace, se kterou porovnááme. Nejčastějším typem redukce je komprese depth map. Těchto metod bylo vyvinuto pouze málo, protože nemá smysl redukce této informace, ale spíše úplné odstranění porovnávání. Z tohoto důvodu zde nebude uveden podrobnější popis metod z této kategorie.

Metody bez reference extrahují z videa parametry, na jejichž základě lze poté zhodnotit jeho kvalitu. Jejich největší předností je možnost určovat kvalitu v online aplikacích [7]. Jejich korelace se subjektivními testy je ale přirozeně menší než u metod s referencí.

1.1 Artefakty a zkreslení typická pro stereoskopické video

V této kapitole se pokusím stručně popsat základní zkreslení, která jsou typická pro stereoskopické video. Většina známých artefaktů a zkreslení je popsána v [8][9][10].

- Accomodation convergence rivalry [8][9][10]

Pozorovatel odhaduje vzdálenost pro blízké předměty primárně na základě dvou ukazatelů. Jedná se o zaostření čočky (accomodation) a binokulární vidění (convergence). Tyto ukazatelé jsou reflexivně spojené. Při sledování stereoskopického videa ale dochází k rozporu mezi těmito ukazateli. Pokud tedy pomocí binokulárního vidění odhadneme pozorovací vzdálenost (vzdálenost od pozorovatele k displayi), také na tuto vzdálenost zaostříme. Stereoskopické video ale navozuje vjem prostoru, tudíž předměty, které se jeví blíže/dále k pozorovateli, budou neostré. Tento jev může způsobovat výrazné nepohodlí, únavu, bolest hlavy, nevolnost, závrať.

- Puppet theatre effect - efekt loutkového divadla [8][9]

Při sledování stereoskopického videa může nastat situace, při které se ukazatele vzdálenosti pomocí binokulárního vidění (binocular cues) dostanou do kolize s ukazateli, na něž je potřeba pouze jedno oko (monocular cues, např velikost, překrývání objektů). Binocular cues mají větší vliv na vnímání vzdálenosti, takže zastíní monocular cues. Pokud bude rozpor veliký, může dojít k tomu, že objekty a postavy v pozorované scéně se budou jevit malé a připomínat loutkové divadlo.

- Card board effect [8][9]

Toto zkreslení vzniká při kompresi a kódování videa. Pokud příliš zkomprimujeme depth map, bude se ve scéně vyskytovat pouze několik hodnot hloubky. To způsobí nepřirozeně plochý vzhled objektů.

- Depth bleeding/ringing [8]

Přílišná komprese depth map vede k rozmazání a rozpití hran. Tento jev je označován depth bleeding nebo ringing.

- Crosstalk [11]

Další jevem, který způsobuje nepohodlí a únavu je tzv. crosstalk. Crosstalk vzniká při nedokonalém oddělení obrazů pro levé a pravé oko při zobrazování. Vizualně se crosstalk projevuje jako tzv. "duchové"

- Sheer distortion [8][9]

Toto zkreslení vzniká také při zobrazování na auto-stereoskopický display. Většina auto-stereoskopických technologií pro zobrazování předpokládá, že se pozorovatel nehýbe. Pokud dojde k pohybu pozorovatele vnímání vzdáleností se zkreslí.

- Cross distortion [8][9][3]

Pokud mají obrazy pro levé a pravé oko rozdílnou kvalitu, subjektivně se výsledná kvalita blíží kvalitě lepšího z obrazů. Při kompresi se tohoto jevu využívá v tzv. asymetrickém kódování, kdy jeden obraz je možné více zkomprimovat. V případě, že rozdíl v kvalitě obrazů bude velký, dojde ke zkreslení vnímání hloubky.

1.2 Přehled používaných formátů a standardů

Tato sekce se bude zabývat nejpoužívanějšími formáty a standardy, které jsou běžné pro stereoskopické video. Při vývoji těchto formátů a standardů je důležité, aby byly univerzální pro všechny typy zobrazovacích technologií [3][12].

- CSV - conventional stereo video [3]

Nejjednodušší reprezentací stereoskopického videa je formát CSV. Stereoskopické video vzniká nasnímáním scény pomocí dvou nebo více kamer. Formát CSV je reprezentován přímo výstupy těchto kamer. V případě dvou kamer se těmito pohledům říká novel views nebo stereopair. Pokud chceme pro přenos tohoto formátu použít stávající protokoly a kodeky, stačí pouze použít časové nebo prostorové prokládání. Hlavní nevýhodou tohoto formátu je nemožnost upravovat video při zobrazování a tím ho přizpůsobit zobrazovací technologii. Vjem hloubky je dán umístěním kamer při snímání a už jej nelze měnit.

- V+D - Video plus Depth [3][12]

Nevýhody CSV částečně odstraňuje formát V+D. Video je reprezentováno jedním referenčním pohledem a depth map, jak můžeme vidět na obrázku 1.1 [15]. Depth map je typicky šedotónový obraz. Hodnota jeho pixelů značí vzdálenost odpovídajícího pixelu v referenčním pohledu od bodu pozorování (viewpoint). Většinou je depth map reprezentován 8 bity. Máme tedy k dispozici 256 úrovní hloubky. Hodnota 255 (bílá) značí nejbližší bod, 0 (černá) nejvzdálenější. Pomocí algoritmu DIBR - depth image based rendering můžeme před zobrazením videa z tohoto formátu vytvořit novel views [13]. Optimalizací tohoto algoritmu můžeme docílit přizpůsobení videa danému zobrazovacímu systému. Největší nevýhodou tohoto přístupu je zkreslení v tzv. occlusion

oblastech [14]. V druhém pohledu, který renderujeme pomocí DIBR, jsou oblasti, jež nejsou v referenčním pohledu viditelné (occlusion oblasti). Nemáme žádnou informaci o těchto oblastech a tak dochází pouze k odhadu, který nutně vede ke zkreslení.



Obrázek 1.1: Obraz a jeho depth map [15]

- MVD - Multiview video plus depth [12][16][17]

Formát MVD odstraňuje přítomnost zkreslení v occlusion oblastech u V+D. Docílíme toho nasnímáním scény několika kamerami, které jsou od sebe posunuté v horizontální rovině. MVD se tedy skládá z několika referenčních pohledů a jim odpovídajících depth map. Protože přenášíme vždy celé pohledy, tudíž i informace o oblastech, ve kterých nedochází ke zkreslení při DIBR, dochází k přenosu poměrně velkého objemu redundantní informace. Tento formát je vhodný pro auto-stereoskopické displaye.

- LDV - Layered depth video [12][17]

LDV je v podstatě zkomprimovaný MVD o redundantní informaci. Skládá se z referenčního pohledu, jemu odpovídající depth map, obrazu s informací o occlusion oblastech a jim odpovídající depth map. Pokud má LDV pouze tyto 4 obrazy, jedná se o tzv. zjednodušené LDV. Stejně jako MVD je LDV formát vhodný pro auto-stereoskopické displaye.

- DES - Depth enhanced stereo [3][17]

Jak už bylo výše zmíněno, klíčovou vlastností u formátů pro stereoskopické video by měla být univerzálnost, proto byl zaveden formát DES, který je kompatibilní s auto-stereoskopickými displayi a stejně tak i se stereoskopickým zobrazením. Video je reprezentováno v podstatě dvojicí zjednodušených LDV, z nichž každé odpovídá jednomu pohledu (levý/pravý).

- FVT - Free view point television [18]

Nejnovějším trendem v oblasti 3D produkce obrazu je FVT. Jak napovídá název, tento formát počítá s volným bodem pozorování. Můžeme tedy zvolit přímo bod, ze kterého budeme scénu pozorovat. Vše je ale ještě ve stádiu vývoje, protože je velice těžké při

snímání reálné scény získat dostatek informací, ze kterých je poté možné renderovat video pro libovolný bod pozorování. Nejčastěji se k tomu používá systému, který využívá 100 kamer.

Kapitola 2

Metody s referencí

Tyto metody jsou založené na porovnání videa, u kterého zkoumáme kvalitu s referenčním videem. Pod pojmem referenční video máme na mysli to samé video před kódováním, kompresí, zkrácením přenosovou cestou atd. Naprostá většina těchto metod vychází z dnes už běžných metod pro posouzení kvality videa. Použitím těchto metod posoudíme pouze kvalitu obrazu, ale nebereme v potaz specifika, která jsou běžná pro stereoskopické video. Je snaha rozšířit tyto metody tak, aby byly schopné postihnout tyto jevy. Dalo by se říci, že struktura metod s referencí je ve většině případů taková, že aplikací známé metody (SSIM, PSNR, VQM...) posoudíme kvalitu obrazu. Dále vhodným algoritmem obsáhneme specifika stereoskopického videa (disparity, crosstalk, vjem hloubky, komfort při sledování, únavu, zkrácení v occlusion oblastech, pohyb...) a výsledné hodnocení je kombinací těchto dílčích hodnocení.

Z důvodu snížení výpočetní náročnosti se metody aplikují pouze na jasovou složku. Pokud je v popisu metody uveden její postup pouze pro obraz, předpokládá se s aritmetickým průměrováním přes snímky videosekvence.

2.1 Metody založené na SSIM - Structural similarity index metric/index strukturální podobnosti

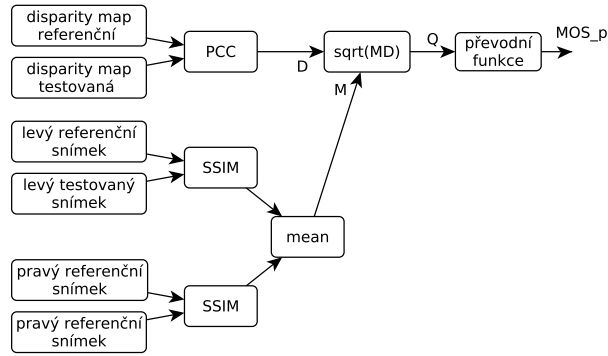
V literatuře [19] je popsána metoda hodnocení kvality na základě strukturálního zkrácení. Bylo dokázáno, že pomocí strukturálního zkrácení lze lépe popsat kvalitu obrazu. To je způsobeno především tím, že lidský vizuální systém je přizpůsoben na extrahování informace o struktuře (strukturální informace) z pozorované scény [19]. Tato metoda má větší korelaci se subjektivními testy, než metody založené na měření chyb (MSE, PSNR). Je to způsobeno tím, že metody založené na měření chyb postihnou všechny degradace v obraze, i když pozorovateli subjektivně nevadí. Podle [19] lze hodnotu SSIM vyčíslit podle následující rovnice.

$$SSIM = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2(\bar{y})^2]} \quad (2.1)$$

kde $x = \{x_i | i = 1, 2, \dots, N\}$ je referenční snímek, $y = \{y_i | i = 1, 2, \dots, N\}$ je testovaný snímek, \bar{x} / \bar{y} je aritmetický průměr x/y , σ_x^2 / σ_y^2 představuje rozptyl x/y . SSIM nabývá hodnot $\langle -1, 1 \rangle$, přičemž hodnota 1 odpovídá nejlepšímu výsledku (x, y stejné). Z důvodu úspory výpočetního výkonu se většinou tento výpočet uplatní na náhodně zvolené bloky v obrazu. Velikost těchto bloků je nejčastěji 8×8 . Výsledné hodnocení obrazu je pak aritmetický průměr hodnocení bloků.

2.1.1 Using disparity for quality assesment of stereoscopic images - 2008

Tato metoda je popsána v literatuře [20]. K SSIM je přidáno hodnocení hloubky na základě informace o disparity. Blokové schéma této metody je na obrázku 2.1 [20].



Obrázek 2.1: Blokové schéma metody [20]

Pod pojmem disparity myslíme vzdálenost mezi odpovídajícími pixely v levém a pravém snímku. Čím větší je tedy tento vektor, tím větší by měl být vjem hloubky [5][20]. Pro výpočet disparity je použit algoritmus popsáný v literatuře [21], ve kterém se předpokládá, že sobě odpovídající pixely mají podobnou jasovou hodnotu. Ke každému stereopáru získáme disparitní obraz. Disparitní obraz má stejnou velikost jako velikost stereopáru a hodnota jeho pixelů je hodnota disparity pro odpovídající pixely. Disparitní obraz referenčního a testovaného stereopáru porovnáme pomocí globálního korelačního koeficientu, protože se nejedná o reálné obrazy (natural image) [20]. Nejčastěji se používá Pearsonův korelační koeficient vyjádřený rovnicí 2.2 [22]. Tímto tedy získáme hodnocení kvality na základě disparity D_d .

$$PCC = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2}} \quad (2.2)$$

Kvalitu obrazu M nejefektivněji určíme jako průměr SSIM indexů pro levý a pravý pohled [20]. Poté byly v [20] vyzkoušeny různé způsoby, jak zkombinovat oba dílčí výsledky do celkového hodnocení kvality Q_f pro daný snímek. Největší korelaci (PCC 0,902) se subjektivními testy má ten vyjádřený v rovnici 2.3 [20].

$$Q_f = \sqrt{MD_d} \quad (2.3)$$

Ukazatel Q_f nabývá hodnot $\langle -1, 1 \rangle$. Pro posouzení korelace se subjektivními testy $DMOS$ musíme Q_f upravit pomocí převodní funkce. Byla použita funkce 2.4 [20], jejíž koeficienty b_1, b_2, b_3 jsou vhodně nastaveny regresní metodou.

$$DMOS_p = \frac{b_1}{(1 + \exp(-b_2 * Q_f - b_3))} \quad (2.4)$$

2.1.2 Quality evaluation of stereoscopic videos using depth map segmentation - 2011

Metoda popsaná v [23] používá rozdělení obrazu na nepřekrývající se vrstvy (popředí/pozadí). Toho docílíme segmentací histogramu depth map na základě algoritmu popsaného v literatuře [24]. Poté vypočítáme SSIM index pro každou vrstvu. U bloků, které leží na hranici, budeme vždy počítat jen s tou částí bloku, která patří do dané vrstvy, aby se zamezilo rušení mezi vrstvami. SSIM index pro danou vrstvu získáme váhovanou sumací podle rovnic 2.5 a 2.6 [23]. L_{ij} se nazývá index kvality vrstvy (layer quality index).

$$L_{ij} = \frac{\sum_{k=0}^{n_{ij}} w_{ijk} SSIM_{ijk}}{\sum_{k=0}^{n_{ij}} w_{ijk}} \quad (2.5)$$

$$w_{ijk} = \begin{cases} 1 + \frac{z_{near} - z_{ij}}{z_{near} - z_{far}} & \text{pro bloky ležící na hranici} \\ 1 & \text{jinde} \end{cases} \quad (2.6)$$

kde $SSIM_{ijk}$ je SSIM index pro k -tý blok, j -tou vrstvu a i -tý snímek, w_{ijk} je váhovací faktor pro k -tý blok, j -tou vrstvu a i -tý snímek a n_{ij} je počet bloků v j -té vrstvě, z_{near} vyjadřuje hodnotu nejbližšího (nejvyšší hodnota) pixelu v depth map, z_{far} vyjadřuje hodnotu nejvzdálenějšího (nejnižší hodnota) pixelu v depth map, z_{ij} je průměr pixelů depth map, které odpovídají dané vrstvě. Je zřejmé, že váhování pomocí w_{ijk} nemá vliv na bloky uvnitř dané vrstvy a zvýší hodnotu bloků ležících na hranici. Poté z jednotlivých indexů kvality vypočítáme index kvality snímku F_i podle rovnice 2.7 [23].

$$F_i = \frac{\sum_{j=0}^{N_i} m_{ij} L_{ij}}{\sum_{j=0}^{n_i} m_{ij}} \quad (2.7)$$

$$m_{ij} = S_{ij} \frac{z_{near} - z_{ij}}{z_{near} - z_{far}} \quad (2.8)$$

kde m_{ijk} je váhovací faktor pro j-tou vrstvu a i-tý snímek a n_{ij} je počet vrstev v j-tém snímku, S_{ij} je počet pixelů v j-té vrstvě a i-tém snímku. Výsledný index kvality pro video sekvenci se vypočítá z rovnic 2.9, 2.10 a 2.11 [23].

$$Q = \frac{\sum_{i=0}^I W_i F_i}{\sum_{i=0}^I W_i} \quad (2.9)$$

$$M_{ij} = \frac{\sum_{k=0}^{n_i} m_{ik}}{K_M} \quad (2.10)$$

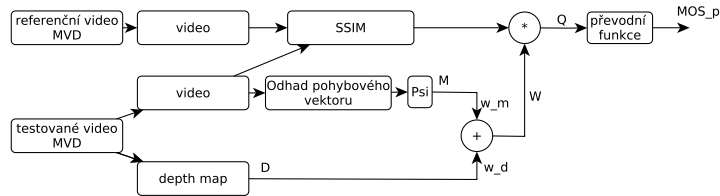
$$W_i = \begin{cases} \sum_{k=0}^{n_i} m_{ik} & \text{pro } M_i \leq 0.8 \\ \frac{(1.2 - M_i)/4}{\sum_{k=0}^{n_i} m_{ik}} & \text{pro } 0.8 \leq M_i \leq 1.2 \\ 0 & \text{pro } M_i \geq 1.2 \end{cases} \quad (2.11)$$

kde W_i je váhovací faktor pro i-tý snímek, I je počet snímků v sekvenci, m_{ik} je velikost pohybového vektoru k-tého bloku, K_M je normalizační konstanta.

Uvedený postup můžeme uplatnit na video ve formátu MVD - Multi-view Video plus Depth, který je reprezentován přímo videem a depth map. V případě práce přímo s levým a pravým pohledem (novel views), je nutné nejdříve z nich vypočítat informaci o hloubce (depth map). V tomto případě byl využit algoritmus DERS - depth estimation reference software, který je založen na porovnávání bloků v novel views [25]. Celý postup můžeme uplatnit zvláště na levý a pravý snímek. Nejeftektivnější celkový výsledek bude průměr těchto hodnocení [5][20].

2.1.3 Depth Perception and Motion Cue based 3D Video Quality Assesment - 2012

(vyšší hodnoty pixelů v depth map) Jak napovídá již název, tato metoda kombinuje SSIM spolu s informací o hloubce a pohybu. Schéma této metody je na obrázku 2.2 [26].



Obrázek 2.2: Blokové schéma metody [26]

Je založená na tom, že oblasti, které se jeví tak, že vystupují z obrazu, na sebe strhávají pozornost [26]. Zkreslení (degradace) v těchto oblastech jsou pozorovatelem vnímány daleko více, než ve zbytku obrazu. Při výskytu většího množství pohybu (vyšší hodnota pohybového vektoru), klesá úroveň vnímání hloubky a subjektivní vnímání kvality [26][27]. Na základě těchto dvou parametrů je poté stanoven váhovací obraz.

Odhad pohybového vektoru

Jedním z nejpoužívanějších algoritmů na odhad pohybového vektoru je block based optical flow algorithm, který popisuje rovnice 2.12 [26]

$$(x_d, y_d) = \arg \min_{x'_d, y'_d \in \{0, \pm 1, \pm 16\}} \sum_{i,j}^W |f_{v,t}(i, j) - f_{v,t-1}(i + x'_d, j + y'_d)| \quad (2.12)$$

kde x_d/y_d je horizontální/vertikální souřadnice pohybového vektoru, $f_{v,t}$ je aktuální snímek a $f_{v,t-1}$ je snímek předchozí, W označuje velikost bloku pro který stanovujeme pohybový vektor. Jeho velikost je v tomto případě 16. Informaci o pohybu $M_{v,t}$ kterou použijeme při dalším výpočtu stanovíme z rovnic 2.13 a 2.14 [26].

$$m_{v,t} = \sqrt{|x_d| + |y_d|} \quad (2.13)$$

$$M_{v,t} = \Psi(m_{v,t}) \quad (2.14)$$

kde Ψ je lineární funkce, která upraví hodnotu tak, aby byla v rozmezí 0-255. Hodnota 255 značí největší množství pohybu.

Váhování na základě depth map

Tato metoda pracuje s videem ve formátu MVD. Depth map máme přímo k dispozici. Pokud máme video v takovém formátu, že není k dispozici depth map, musíme jí nejdříve odhadnout. K tomuto odhadu se běžně používá algoritmus DERS [25].

Na základě výše pospaných parametrů stanovíme váhovací obraz dle rovnice 2.15 [26].

$$W = w_m \cdot M + w_d \cdot D \quad (2.15)$$

Protože při subjektivních testech se prokázalo, že pohyb je vnímán více než artefakty související s hloubkou, je $w_m = 0.65$ a $w_d = 0.35$.

Na porovnání výsledků metody se subjektivními testy bylo použito pouze dvou sekvencí (*cafeapoznanhall2*, 1920×1080 , 10s). Subjektivních testů se zúčastnilo 24 pozorovatelů (věk 19-25, 22 mužů a 2 ženy, 22 z nich s touto problematikou nepřišlo nikdy do styku). Při

porovnání se ukázalo, že tato metoda překonává klasické SSIM, především pokud je použito asymetrické kódování [26].

2.1.4 Assessment of Stereoscopic Crosstalk Perception

Jak napovídá název, metoda popsaná v [28], je navržena tak, aby byla schopna postihnout jev zvaný crosstalk. V důsledku nedokonalé separace pohledů lze jedním okem vidět nejen pohled pro ně určený, ale také část druhého pohledu. Tento jev nazýváme crosstalk a dochází k němu ve většině dnes dostupných 3D displayů. Projevuje se hlavně tím, že jsou vidět tzv. "duchové" [11]. Nejvíce se crosstalk objevuje podél hran, nebo v místech, kde je velký kontrast [28].

Nejdříve vypočítáme SSIM mapu, což je obraz, jehož pixely nabývají hodnot $\langle 0, 1 \rangle$. Přičemž 1 značí, že obrazy jsou stejné. Tuto mapu získáme použitím SSIM algoritmu na celé obrazy a následně nanormujeme, protože SSIM nabývá hodnot $\langle -1, 1 \rangle$. Získání SSIM mapy L_s je vyjádřeno v rovnici 2.16

$$L_s = \text{SSIM}(L_0, L_c) \quad (2.16)$$

kde L_0/L_c je původní/degradovaný obraz a *SSIM* značí použití SSIM algoritmu. Následně vezmeme v potaz informaci o hloubce, protože pokud se crosstalk objeví v oblastech, které se mají jevit blíže pozorovateli, bude vnímán více [26][28]. Ke zpřesnění odhadu depth map použijeme DERS algoritmus [25]. Zpřesněnou depth map tedy získáme podle rovnice 2.17 [28].

$$R_{dep} = \text{DERS}(R_0) \quad (2.17)$$

kde R_0/R_{dep} je původní/zpřesněná depth map a *DERS* značí použití DERS algoritmu. Subjektivními testy bylo zjištěno, že pokud je hodnota SSIM vyšší než 0.977, není crosstalk viditelný [28]. Na základě tohoto poznatku upravíme R_{dep} podle 2.18 [28].

$$R_{pdep}(i, j) = \begin{cases} R_{pdep}(i, j) & \text{pro } SSIM(i, j) < 0.977 \\ 0 & \text{pro } SSIM(i, j) \geq 0.977 \end{cases} \quad (2.18)$$

kde i, j jsou indexy pixelu. Následně bylo v [28] vyzkoušeno několik způsobů jak zkombinovat L_s a R_{pdep} . Tady uvedeme pouze způsob s nejvyšší korelací se subjektivními výsledky. Celkové hodnocení kvality Q je vyjádřeno rovnicemi 2.19 2.20 [28].

$$C_{pdep} = L_s \times \left(\frac{1 - R_{pdep}}{255} \right) \quad (2.19)$$

$$Q = \text{AVG}(C_{pdep}) \quad (2.20)$$

Subjektivních testů, které byly posléze použity na hodnocení metody se zúčastnilo 28 osob (věk 23-46, 15 mužů a 13 žen). Pro tyto testy bylo použito 6 sekvencí s různým uměle vytvořeným crosstalkem (0, 5%, 10% a 15%) a s různou vzdáleností snímacích kamer (baseline: 0, 4, a 12 cm). Celkem tedy 72 sekvencí. Při porovnání se subjektivními výsledky byl korelační koeficient $PCC = 0,884$, což je výrazně lepší, než při použití samotného SSIM ($PCC = 0,825$).

2.2 Metody založené na MSE - Mean square error

Tyto metody používají jako hlavní nástroj na porovnání střední kvadratickou odchylku - MSE. Bylo dokázáno, že použití MSE není vhodné pro tuto aplikaci [19][26]. Nicméně v literatuře [5] byla v roce 2011 publikována metoda, která je založená na MSE a 3D-DCT. Tuto metodu uvedeme jako jediného zástupce v této kategorii, protože dosahuje dobrých výsledků.

2.2.1 3D-DCT based perceptual quality assessment of stereo video 2011

Tato metoda pracuje s levým a pravým pohledem (novel views, stereopár). Nejdříve stanovíme disparitní obraz na základě podobnosti bloků přičemž uvažujeme pouze posun v horizontálním směru, který poté váhujeme na základě barevné difference [5].

Velikost okna byla zvolena 4x4. Vezmeme tedy referenční blok z levého snímku A_0 , k němu najdeme nejvíce podobný blok v tom samém snímku A_1 . V pravém snímku vezmeme jako referenční blok A_d blok na stejné pozici jako A_0 posunutý o odpovídající hodnotu disparity. K němu najdeme dva nejvíce podobné bloky z pravého snímku A_2, A_3 . Tyto podobné bloky nehledáme v celém snímku, ale pouze v tzv. hledacím prostoru. V tomto případě bylo zvoleno 22×22 . Z těchto bloků vytvoříme 3D strukturu $A = \{A_0, A_1, A_2, A_3\}$, kterou transformujeme podle 3D-DCT. Stejným způsobem získáme $B = \{B_0, B_1, B_2, B_3\}$ z referenčního videa. Použitím DCT modelujeme dva jevy, k nimž dochází při pozorování. Jedná se o binokulární vidění a pseudonáhodné pohyby očí (saccades).

Získali jsme struktury $4 \times 4 \times 4$ koeficientů 3D-DCT, jež porovnáme pomocí modifikovaného MSE vyjádřeného rovnicí 2.21 [5].

$$MSE_{3D} = \frac{16}{I \times J} \sum_{i=1}^{I-3} \sum_{j=1}^{J-3} MSE(A_{ij}^D, B_{ij}^D) \cdot C_{4 \times 4}^2 \cdot Mask \quad (2.21)$$

kde I, J je velikost obrazu, A_{ij}^D značí první vrstvu z 4x4x4 3D-DCT koeficientů na pozici i, j z referenčního obrazu, B_{ij}^D značí první vrstvu z 4x4x4 3D-DCT koeficientů na pozici i, j

z testovaného obrazu, $C_{4 \times 4}^2$ je kvadrát normalizační kvantizační tabulky JPEG přepočítaný (downsample) na velikost 4×4 [31], *Mask* slouží k redukci kontrastního maskování mezi koeficienty. Výsledný index kvality určíme z rovnice 2.22 [5]

$$PHVS_{3D} = 10 \log \left(\frac{255^2}{MSE_{3D}} \right) \quad (2.22)$$

Subjektivních testů použitých na porovnání s výsledky metody se zúčastnilo 32 osob (věk 18-45, 50% žen). Byl vybrány 4 sekvence, u nichž bylo možné měnit baseline(2D, short, long) a kvantizační parametry komprese MPEG-4 (25, 30, 35, 40, 45). Celkově bylo 12 referenčních, nezkreslených a 60 zkreslených sekvencí. Tato metoda měla PCC 0,9063.

Kapitola 3

Metody bez reference

Jak napovídá název, tyto metody nevyužívají porovnání hodnoceného videa s referenčním videem. Jsou tedy vhodné zejména k aplikacím v reálném čase [7]. Obecný princip těchto metod se dá shrnout do dvou kroků. První krok je výpočet parametrů, které korelují se subjektivními testy. Druhým krokem je kombinace parametrů do výsledného hodnocení kvality. Kombinace spočívá ve výběru vhodné regresní metody a nastavení jejích koeficientů na základě korelace s testovacími daty. Pod pojmem testovací data se rozumí část databáze, která je k tomuto nastavení koeficientů použita. Běžně se jako testovací data používá kolem 60% náhodně vybraných videí z databáze. Korelace se subjektivními testy je pro metody bez reference přirozeně nižší než u metod s referencí.

3.1 Stereoscopic video quality perception model - SV-QPM 2011

Metoda SV-QPM je podrobně popsána v literatuře [27]. Kvalita videa se určí z těchto parametrů: časová změna pohybového vektoru (temporal variance), variace hodnot disparity v rámci snímku (disparity variation in intra-frames), mezisnímková variace hodnot disparity (disparity variation in inter-frames), rozložení disparity v hraničních oblastech snímku (disparity distribution of frame boundary areas). Tyto parametry jsou zkombinovány do celkového hodnocení lineární regresí.

3.1.1 Disparity

Protože nemáme k dispozici depth map, musíme ji nejdříve odhadnout. K tomu poslouží algoritmus založený na podobnosti bloků (horizontal block matching algorithm) popsáný v literatuře [33]. Hodnota disparity se určí podle rovnice 3.1, kde x, y jsou souřadnice bloku, i je posun v horizontálním směru a SR (search range) udává rozsah, ve kterém určujeme

disparity. V tomto případě je rozsah ± 31 . $MSE^n(x, y, i)$ je střední kvadratická odchylka mezi blokem na souřadnicích (x, y) v levém a pravém obrazu a vyjádřit ji můžeme pomocí rovnice 3.2 [27], kde $f_L^n(x, y), f_R^n(x, y)$ je hodnota intenzity pixelu na souřadnicích (x, y) v n-tém levém, respektive pravém obrazu a kde M, N je výška respektive šířka bloku.

$$disp^n(x, y) = \arg \max_{i \in SR} MSE^n(x, y, i) \quad (3.1)$$

$$MSE^n(x, y, i) = \frac{1}{M \times N} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \{f_L^n(x+p, y+q) - f_R^n(x+p, y+q)\}^2 \quad (3.2)$$

Hodnota disparity je zpravidla v rozsahu 0-255, proto odhadnutou hodnotu normujeme dle 3.3 [27].

$$f_{disparity}^n(x, y) = disp^n \times \frac{255}{DB} + 128 \quad (3.3)$$

kde DB (depth budget) určuje počet možných hodnot disparit. V tomto případě je tato hodnota 63 (rozsah ± 31). Následně rozdělíme každý disparitní obraz $f_{disparity}^n(x, y)$ na bloky 32×32 a spočítáme podle rovnice 3.4 průměrnou hodnotu pro každý blok.

$$f_{avg}^n(x, y) = \frac{1}{W \times H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} f_{disparity}^n(i, j) \quad (3.4)$$

kde W, H je výška, respektive šířka bloku a $f_{disparity}^n(i, j)$ je hodnota pixelu disparitního obrazu o souřadnicích (i, j) .

3.1.2 Časová změna pohybového vektoru

Je všeobecně známo, že větší množství pohybu zhoršuje vjem hloubky. To způsobí horší subjektivní kvalitu. Nejdříve odhadneme pohybový vektor. Můžeme použít například výše zmíněný block based optical flow algorithm [26] a vezmeme v potaz jeho průměrnou velikost vyjádřenou rovnicí 3.5 [27].

$$tv = \frac{1}{N \times B} \sum_{n=0}^{N-1} \sum_{b=0}^{B-1} \sqrt{(MV_{b,x}^n)^2 + (MV_{b,y}^n)^2} \quad (3.5)$$

kde n je číslo snímku, N je celkový počet snímků, b je číslo bloku ve snímku, B je celkový počet bloků, $MV_{b,x}^n, MV_{b,y}^n$ jsou hodnoty x,y-ové souřadnice pohybového vektoru MV_b^n pro n-tý blok a b-tý snímek. Následně výsledné hodnoty roztrídíme do pěti kategorií podle rovnice 3.6 [27].

$$TV = \begin{cases} 5 & \text{pro } TV > 4.5 \\ 4 & \text{pro } 3.5 < TV < 4.5 \\ 3 & \text{pro } 2.5 < TV < 4.5 \\ 2 & \text{pro } 1.5 < TV < 4.5 \\ 1 & \text{pro } TV \leq 1.5 \end{cases} \quad (3.6)$$

3.1.3 Variace hodnot disparity v rámci snímku (Intra-frame disparity variation)

Pokud dochází k velkým změnám hodnoty disparity v rámci jednoho snímku, subjektivní kvalita klesá. Tuto skutečnost do metody zahrneme tak, že použijeme $f_{avg}^n(x, y)$ z rovnice 3.4 a vypočítáme variaci hodnot disparity (intra-frame disparity variation) pro blok o souřadnicích (p, q) $DV_s(p, q)$ dle 3.7 [27]

$$DV_s(p, q) = \frac{1}{8} \sum_{i=-1}^1 \sum_{j=-1}^1 \left| f_{avg}^n(p, q) - f_{avg}^n(p+i, q+j) \right| \quad (3.7)$$

Následně vypočítáme průměrnou variaci hodnot disparity v rámci snímku DV_s podle 3.8 [27].

$$DV_s = \frac{1}{N} \sum_{n=0}^{N-1} \sqrt{\frac{1}{K \times L} \sum_{p=0}^{K-1} \sum_{q=0}^{L-1} \{DV_s^n(p, q)\}^2} \quad (3.8)$$

kde K, L je počet sub-snímku (sub-image) v řádcích a sloupcích. Byla použita hodnota 32. N je celkový počet snímků.

3.1.4 Mezisnímková variace hodnot disparity (Inter-frame disparity variation)

Pokud dochází k velkým změnám hodnoty disparity mezi jednotlivými snímky, klesá subjektivní kvalita. Proto vezmeme v úvahu mezisnímkovou variaci hodnot disparity DV_t vyjádřenou rovnicí 3.9 [27].

$$DV_t = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{K \times L} \sum_{p=0}^{K-1} \sum_{q=0}^{L-1} \left| f_{avg}^n(p, q) - f_{avg}^{n-1}(p, q) \right| \quad (3.9)$$

3.1.5 Rozložení disparity v hraničních oblastech snímku (Disparity distribution of frame boundary area)

Pokud se v hraničních oblastech vyskytují malé nebo naopak velké hodnoty disparity, dochází ke kolizi odhadu vzdálenosti na základě paralaxy (motion parallax depth cue) a occlusion depth cue (objekt, který částečně zakrývá jiný objekt, je blíže [1]). Tento jev výrazně snižuje komfort při sledování a může způsobit např. nevolnost nebo bolest a motání hlavy [8][9].

Nejdříve určíme hraniční oblast snímku. Její šířka, výška (bráno od kraje snímku) je osmina šířky, respektive výšky snímku. V této oblasti vypočítáme rozložení disparity D_b podle 3.10 [27].

$$D_b = \frac{1}{N} \sum_{n=0}^{N-1} \sqrt{\frac{1}{BC} \sum_{i \in B} \sum_{j \in C} \{f_{avg}^n(p, q) - 128\}} \quad (3.10)$$

kde BC je počet bloků v hraniční oblasti, B, C značí hraniční oblast. Odečítáním hodnoty -128 docílíme toho, že rozložení disparity bude vztaženo k nulové hodnotě.

3.1.6 Stereoscopic video quality perception model (SV-QPM)

Nyní je potřeba všechny dílčí parametry (TV, DV_s, DV_t, D_b) sloučit do celkového hodnocení kvality pomocí lineární regrese. $SV - QPM$ lze vyjádřit rovnicí 3.11 [27].

$$SV - QPM = w_0 + w_1 \times \log TV + w_2 \times DV_s^2 + w_3 \times \sqrt{DV_s} + w_4 \times \sqrt{\frac{DV_t}{TV}} + w_5 \times D_b + w_6 D_b^2 \quad (3.11)$$

kde w_0, w_1, \dots, w_6 jsou regresní parametry.

Na posouzení efektivity byla použita databáze 105 testovacích video sekvencí v HD rozlišení (1920×1080) a délkou trvání 10s. Z databáze bylo náhodně vybráno 60 sekvencí, které posloužily pro odhad regresních parametrů. Zbýlých 45 sekvencí bylo použito pro vyjádření korelace mezi subjektivními hodnoceními a výsledky metody. Metoda dosáhla hodnoty PCC 0,808.

3.2 Algorithmic assesment of 3D quality of experience for images and videos - 2011

Metoda popsaná v literatuře [6] je založena na předpokladu, že stereoskopické obrazy/video, která působí přirozeně, mají určité rozložení statistických parametrů. Odchýlení od tohoto rozdělení způsobuje nepřirozené vnímání obrazu/video a tudíž subjektivně nižší kvalitu. Je

provedena analýza statistických parametrů disparity map D , prostorové aktivity S (spatial activity/features) a pohybově kompenzované diferenční disparity. Z těchto parametrů je poté stanoveno hodnocení kvality.

3.2.1 Disparity map a její statistické parametry

Nejdříve odhadneme disparity map D . K tomu použijeme algoritmus popsáný v literatuře [29], který jako všechny algoritmy založené na stereo matching předpokládá, že pixely, jež jsou v pravém a levém obrazu posunuty o disparity, mají podobnou jasovou hodnotu. Odhad disparity neprobíhá na úrovni pixelů, ale obraz nejdříve segmentujeme na oblasti, které mají stejnou barvu (homogeneous color). V těchto oblastech předpokládáme jen pomalé změny disparity u jednotlivých pixelů. Pro tyto oblasti vypočítáme odhad disparity a následně upřesníme tento odhad a analyzujeme optimální segmentaci.

Nyní máme k dispozici disparity map D , ze které spočítáme následující statistické parametry

- střední hodnota - $\mu_D = E[D]$
- medián - $med_D = median(D)$
- směrodatná odchylka - $\sigma_D = \sqrt{E[D]}$
- koeficient špičatosti - $\kappa_D = \frac{E[(D-\mu)^4]}{(E[(D-\mu)^2])^2}$
- koeficient šikmosti - $skew_D = \frac{E[(D-\mu)^3]}{(E[(D-\mu)^2])^{\frac{3}{2}}}$
- střední hodnotu diferenční D (mean differential disparity) - $\mu_D = E[\delta D]$
- směrodatná odchylka diferenční D (differential disparity standard deviation) - $\sigma_D = \sqrt{E[(\delta D - \mu_d)^2]}$
- koeficient špičatosti diferenční D - $\kappa_D = \frac{E[(\delta D - \mu_d)^4]}{(E[(\delta D - \mu_d)^2])^2}$
- koeficient šikmosti diferenční D - $skew_D = \frac{E[(\delta D - \mu_d)^3]}{(E[(\delta D - \mu_d)^2])^{\frac{3}{2}}}$

kde δD značí uplatnění Laplaceova operátoru na disparity map.

3.2.2 Spatial activity a její statistické parametry

Spatial activity S získáme pro levý i pravý obraz. Z těchto obrazů vypočítáme gradient a poté vyčíslíme jeho rozptyl pro nepřekrývající se bloky 8×8 . Pro takto získanou S vyjádříme následující statistické parametry.

- střední hodnota - $\mu_S = E[S]$
- koeficient špičatosti - $\kappa_S = \frac{E[(S-\mu_s)^4]}{(E[(S-\mu_s)^2])^2}$
- koeficient šikmosti - $skew_S = \frac{E[(S-\mu_s)^3]}{(E[(S-\mu_s)^2])^{\frac{3}{2}}}$

3.2.3 Diferenční disparity map a její statistické parametry

Již dříve bylo zmíněno, že množství pohybu hraje důležitou roli při vnímání kvality stereoskopického videa [27][34]. Odhad pohybových vektorů je proveden algoritmem ARPS - adaptive rood pattern search. Tento algoritmus využívá standardní blokový přístup, ale pro zrychlení celého výpočtu vhodně upravuje velikost bloku, ve které hledáme pohybový vektor [30]. Pohybové vektory odhadneme pro nepřekrývající se bloky 8×8 . Nyní můžeme vyjádřit rozdíl mezi disparity daného bloku v i -tém snímku a jemu odpovídajícímu (pohybově kompenzovanému) bloku v $(i-1)$ snímku. Takto vznikne diferenční disparity map DD a pro ní vyjádříme následující statistické parametry.

- variační koeficient - $VC = \frac{\sqrt{E[DD]}}{E[DD]}$
- medián - $med = median(DD)$
- směrodatná odchylka DD - $\sigma = \sqrt{E[DD]}$
- koeficient špičatosti - $\kappa_s = \frac{E[(DD-\mu_{dd})^4]}{(E[(DD-\mu_{dd})^2])^2}$
- koeficient šikmosti - $skew = \frac{E[(DD-\mu_{dd})^3]}{(E[(DD-\mu_{dd})^2])^{\frac{3}{2}}}$

Nyní máme pro každý snímek 25 statistických parametrů. To je poměrně velké množství, a protože byla použita databáze s poměrně málo videi, bylo nutné zredukovat jejich počet. K tomu bylo použito Analýzy hlavních komponent (PCA). Nakonec ještě upravíme vybrané parametry lineární regrese tak, aby odpovídaly subjektivnímu hodnocení kvality (MOS), dle 3.12 [6].

$$y = \vec{x}^T \vec{\gamma} + \eta \quad (3.12)$$

kde \vec{x} je vektor vybraných parametrů, $\vec{\gamma}$ je vektor váhovacích parametrů, a η je konstanta.

K dispozici je databáze 6 sekvencí snímaných z 5 různých vzdáleností, celkem tedy 6×5 videí. Sekvence jsou v rozlišení 1920×1080 , mají 25 fps a trvají 10s. Z této databáze použijeme polovinu sekvencí 3×5 na natrénování analýzy hlavních komponent a parametrů lineární regrese $(\vec{\gamma}, \eta)$. Ze zbylých sekvencí je 1×5 použito na ověření správnosti předchozích kroků a zbytek, tedy 2×5 , na otestování metody. K posouzení korelace se subjektivními výsledky byl použit Spearmanův koeficient. Jeho aritmetický průměr je 0,76 a variace 0,25.

3.3 A no-reference quality measure for DIBR-based 3D videos - 2011

Jak již název napovídá, metoda, která byla publikována v literatuře [13], slouží k hodnocení kvality videa vzniklého renderováním pomocí depth image based rendering - DIBR. Je zaměřena na ta zkreslení a degradace, která jsou typická pro DIBR.

Nejdříve odhadneme ideální depth map. Renderovaný obraz pomocí této depth map bude bez zkreslení. Následně tuto depth map použijeme pro vypočítání parametrů, které charakterizují zkreslení (spatial outliers - SO, temporal outliers - TO, temporal inconsistencies - TI). Tyto parametry jsou do celkového hodnocení kvality sloučeny pomocí lineární regrese.

3.3.1 Odhad ideální depth map

Při odhadu ideální depth map vyjdeme z rovnice 3.13 [13] popisující wrapping při DIBR

$$\bar{X}_v = \bar{X}_r + s \frac{F_v B}{\bar{Z}} + h \quad (3.13)$$

kde \bar{X}_v je vektor horizontálních souřadnic (horizontal coordinates) virtuální kamery, \bar{X}_r je vektor horizontálních souřadnic referenční kamery, $s = -1/s = +1$ pokud odhadujeme levý/pravý pohled, F_v je ohnisko virtuální kamery (většinou stejné jako ohnisko referenční kamery), \bar{Z} je vektor hodnot depth map odpovídající pixelu o souřadnicích (x_r, y_r) a h je horizontální posun os kamer (vzdálenost mezi bodem na ose kamery ve vzdálenosti ohniska a spojnice bodu konvergence kamer a jejího ref. bodu).

Pomocí ideální depth map \bar{Z}_{ideal} lze tedy vyjádřit rovnicí 3.14 [13] vektor horizontálních souřadnic \bar{X}_0 , který odpovídá nezakreslenému pohledu.

$$\bar{X}_0 = \bar{X}_r + s \frac{F_v B}{\bar{Z}_{ideal}} + h \quad (3.14)$$

Odečtením rovnic 3.13, 3.14 a úpravou získáme rovnici 3.15 [13].

$$\bar{Z}_{ideal} = \frac{s F_v}{(\bar{X}_0 - \bar{X}_v) + s \frac{F_v B}{\bar{Z}}} \quad (3.15)$$

Protože pro malý horizontální posun $\Delta \bar{X} = (\bar{X}_0 - \bar{X}_r)$, horizontální posun pro jednotlivé pixely lze odhadnout na základě změny hodnoty jasu pixelu $\Delta I \approx \alpha \bar{X}$ kde α je konstanta. Za tohoto předpokladu lze rovnici 3.15 upravit na rovnici 3.16 [13].

$$\bar{Z}_{ideal} \approx \frac{s F_v}{\alpha (\bar{I}_0 - \bar{I}_v) + s \frac{F_v B}{\bar{Z}}} \quad (3.16)$$

Tato metoda ale nepoužívá referenci, neznáme tedy \bar{I}_0 . Místo $(\bar{I}_0 - \bar{I}_v)$ použijeme funkci $f(\bar{I}_v, \bar{I}_r)$, kde \bar{I}_v je vektor hodnot intenzity z renderovaného obrazu a \bar{I}_r je vektor hodnot

intenzity z referenčního obrazu. Ideální depth mapu vyjádříme tedy jako funkci $f(\bar{I}_v, \bar{I}_r)$ 3.17 [13].

$$\bar{Z}_{ideal} \approx \frac{sF_v}{\alpha(f(\bar{I}_v, \bar{I}_r)) + s\frac{F_v B}{Z}} \quad (3.17)$$

Funkce $f(\bar{I}_v, \bar{I}_r)$ definujeme jako rozdíl intenzity mezi sobě odpovídajícími bloky v referenčním obrazu \bar{I}_r a jemu korespondujícímu bloku v renderovaném obrazu \bar{I}_v . Zvolená hodnota velikosti bloku d ovlivňuje množství šumu v odhadnuté ideální depth map. Čím větší d tím větší množství šumu.

3.3.2 Parametry charakterizující zkreslení

Protože parametry charakterizující zkreslení jsou funkce $\Delta\mathbf{Z}$ definujeme tento parametr v rovnici 3.18 [13].

$$\Delta\mathbf{Z} = |\mathbf{Z}_{ideal} - \mathbf{Z}| \quad (3.18)$$

kde \mathbf{Z} je matice skládající se z vektorů \bar{Z} .

- Spatial outliers - SO

Parametr spatial outliers je definován v rovnici 3.19 [13].

$$\mathbf{SO} = \text{STD}(\Delta\mathbf{Z}) \quad (3.19)$$

kde STD značí směrodatnou odchylku.

- Temporal outliers - TO

Parametr temporal outliers je definován v rovnici 3.20 [13] jako směrodatná odchylka $\Delta\mathbf{Z}$ dvou po sobě jdoucích snímků.

$$\mathbf{TO} = \text{STD}(\Delta\mathbf{Z}_{t+1} - \Delta\mathbf{Z}_t) \quad (3.20)$$

- Temporal inconsistencies (TI)

Parametr TI charakterizuje vizuální diskomfort způsobený rychlou časovou změnou hodnot disparity. Je definován v rovnici 3.21 [13].

$$\mathbf{TI} = \text{STD}(\mathbf{Z}_{t+1} - \mathbf{Z}_t) \quad (3.21)$$

3.3.3 Hodnocení kvality - NR-3VQM

Výše uvedené parametry je nutné zkombinovat do jednoho hodnocení kvality. To je uvedeno v následující rovnici [13].

$$\text{NR} - 3\text{VQM} = K(1 - \text{SO}(\text{SO} \cap \text{TO}))^a(1 - \text{TI})^b(1 - \text{TO})^c \quad (3.22)$$

a, b, c jsou konstanty, které byly stanoveny použitím metody na třech trénovacích sekvencích. Jejich hodnota byla určena takto: $a = 8, b = 8, c = 6$. K je konstanta jejíž pomocí lze NR-3VQM normovat.

Subjektivních testů, které posloužily na zhodnocení efektivity NR-3VQM, se zúčastnilo 20 dobrovolníků, většinou inženýrů s malou zkušeností v daném oboru. Bylo použito 31 video sekvencí délky 30s.

Jak již bylo výše zmíněno, výsledky metody budou záviset na zvolené velikosti bloku d při odhadu ideální depth map. Výsledky (PCC) jsou shrnuty v tabulce 3.1 [13].

$NR - 3VQM$	PCC
$d = 2$	0,8529
$d = 5$	0,8662
$d = 10$	0.8762
$d = 100$	0,9157

Tabulka 3.1: PCC - NR-3VQM

3.4 Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction 2010

Tato metoda je podrobně popsána v literatuře [35]. Je navržena pro videa používající kompresní standard MPEG-2. Každý levý i pravý snímek je neříve rozdělen na bloky o velikosti 8×8 v těchto blocích je vyhodnoceno, jestli v nich jsou hrany (edge) nebo se jedná o bloky bez hran. Následně jsou vypočítány 3 parametry: Artifacts measure - měření artefaktů, disparity feature a temporal measure. Při měření artefaktů je metoda zaměřena na zero-crossing a blockiness, což jsou artefakty spojené se standardem MPEG-2. Disparity a temporal measure mají za úkol postihnout vjem hloubky a pohybu. Tyto parametry jsou nakonec sloučeny do celkového hodnocení kvality tak, že dochází k průběžnému hodnocení kvality. Kvalita se tedy vyhodnotí vždy pro temporal segment (20 snímků).

3.4.1 Segmentace na edge/no-edge bloky

Prvním krokem je segmentace snímků na bloky o velikosti 8x8, ve kterých vyhodnotíme jestli tento blok obsahuje hrany. Bylo k tomu použito algoritmu z [36][37]. Nejdříve vypočítáme pro každý pixel hodnoty směrodatné odchyly pro jeho okolí a velikosti 3x3 a 5x5. Tyto obrazy od sebe poté odečteme 3.23 [37].

$$D_a(m, n) = |STD_{3x3}(m, n) - STD_{5x5}(m, n)| \quad (3.23)$$

kde $m = 1, 2, \dots, M$, $n = 1, 2, \dots, N$, M, N značí velikost snímku. Směrodatná odchylnka takto vzniklých obrazů je dána rovnicí 3.24 [37]

$$D = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M \times N} (\overline{D_a} - D_{ai})^2} \quad (3.24)$$

kde D_{ai} je hodnota D_a pro i -tý pixel a $\overline{D_a}$ je aritmetický průměr D_a . Podle rovnice 3.25 [37] určíme, zda daný pixel náleží do oblasti obsahující hranu.

$$P = \begin{cases} 1 & \text{pro } D_a(m, n) \geq D \\ 0 & \text{jindy} \end{cases} \quad (3.25)$$

kde "1" a "0" indikují edge a no-edge pixely. Rovnice 3.26 [37] určí, zda blok 8x8 je edge/no-edge.

$$if \left(\frac{n_e}{Sum} > th_n \right) \quad (3.26)$$

kde n_e je počet edge pixelů, Sum je počet pixelů v bloku a $th_n = 0.13$ je prahová úroveň stanovená empiricky. Pokud platí tato podmínka, tento blok obsahuje hranu.

3.4.2 Artifacts measure - měření artefaktů

Jak už bylo výše zmíněno v tomto kroku budou zhodnoceny dva artefakty (zero-crossing a blockiness) zvlášť pro levý a pravý snímek a zvlášť pro edge a no-edge oblasti v těchto snímcích. Nejprve je definována diference snímku v horizontálním směru 3.27 [37].

$$d_h(m, n) = I(m, n + 1) - I(m, n) \quad (3.27)$$

kde $I(m, n)$ je jasová hodnota pixelu o souřadnicích (m, n) , $n \in [1, N - 1]$ a $m \in [1, N]$. Blockiness v horizontálním směru pro blok 8x8 je definováno rovnicí 3.28 [37].

$$B = bh = \frac{1}{8} \sum_{i=1}^8 |d_h(i, 8j)| \quad (3.28)$$

Při výpočtu zero-crossing v horizontálním směru nejprve definujeme indikátor znaménka $d_{h-sign}(m, n)$ 3.29[37].

$$d_{h-sign}(m, n) = \begin{cases} 1 & \text{pro } d_h(m, n) > 0 \\ -1 & \text{pro } d_h(m, n) < 0 \\ 0 & \text{jindy} \end{cases} \quad (3.29)$$

Vynásobením dvou po sobě jdoucích řádků $d_{h-sign}(m, n)$ získáme $d_{h-mul}(m, n)$ 3.30 [37].

$$d_{h-mul}(m, n) = d_{h-sign}(m, n) \times d_{h-sign}(m, n + 1) \quad (3.30)$$

d_{h-mul} je definováno pro $n \in [M, N - 2]$. Zero-crossing v horizontálním směru vypočteme podle rovnice 3.31 [37].

$$z_h(m, n) = \begin{cases} 1 & \text{pro } d_{h-mul} < 0 \\ 0 & \text{jindy} \end{cases} \quad (3.31)$$

kde "1" indikuje zero-crossing. Zero-crossing v horizontálním směru pro blok 8x8 vypočteme podle rovnice 3.32 [37].

$$ZC_{bh} = \sum_{i=1}^8 \sum_{j=1}^8 z_h(i, j) \quad (3.32)$$

Diference snímku ve vertikálním směru je definována pomocí rovnice 3.33 [37].

$$d_h(m, n) = I(m + 1, n) - I(m, n) \quad (3.33)$$

kde $I(m, n)$ je jasová hodnota pixelu o souřadnicích (m, n) , $n \in [1, N]$ a $m \in [1, N - 1]$. Pokud uplatníme stejný postup(3.28 - 3.33), dostaneme hodnoty B_{bv} a ZC_{bv} . Blockiness a zero-crossing pro blok definuje rovnice 3.34 [37].

$$B_b = \frac{B_{bh} + B_{bv}}{2}, \quad ZC_b = \frac{ZC_{bh} + ZC_{bv}}{2}, \quad (3.34)$$

Dále vyjádříme průměrnou hodnotu blockiness a zero-crossing pro edge(non uniform) a no-edge(uniform) oblasti v rámci snímku 3.34 [37].

$$Bl_e = \frac{1}{N_e} \sum_{i=1}^{N_e} B_{be}, \quad Bl_n = \frac{1}{N_n} \sum_{i=1}^{N_n} B_{bn} \quad (3.35)$$

kde N_e a N_n je počet edge a no-edge bloku ve snímku. Stejným způsobem vypočítáme Br_n a Br_e pro pravý snímek. Stejný postup uplatníme také v případě zero-crossing 3.34 [37].

$$ZCl_e = \frac{1}{N_e} \sum_{i=1}^{N_e} ZC_{be}, \quad ZCl_n = \frac{1}{N_n} \sum_{i=1}^{N_n} ZC_{bn} \quad (3.36)$$

Nyní přejdeme k operacím v časové oblasti. V té je definován sub-segment, který zahrnuje 4 po sobě jdoucí snímky a temporal segment, který zahrnuje 20 po sobě jdoucích snímků (5 sub-segmentů). Nyní vypočteme tyto artefakty nejdříve pro sub-segment a poté pro temporal segment. Tyto operace jsou shrnuty v rovnicích 3.37 a 3.38 [35].

$$B_{es} = \frac{1}{F} \sum_{f=1}^F BC_{ef}, \quad ZC_{es} = \frac{1}{F} \sum_{f=1}^F CZ_{ef} \quad (3.37)$$

$$B_{et} = \frac{1}{S} \sum_{s=1}^S BC_{es}, \quad ZC_{es} = \frac{1}{S} \sum_{s=1}^S CZ_{es} \quad (3.38)$$

kde $F = 4$ je počet snímků v sub-segmentu a $S = 5$ je počet sub-segmentů v temporal segmentu.

3.4.3 Disparity measure

Pro výpočet disparity byl použit algoritmus, který stanoví pro blok 8x8 hodnotu disparity v rozsahu ± 32 na základě minimálního rozdílu parametru zero-crossing. Na základě takto vypočtené disparity určíme ke každému bodu ZCl_b jemu odpovídající (kompenzovaný) ZCr_b . Diferenci mezi těmito parametry vyjadřuje následující rovnice.

$$DZC_b = ZCl_b \oplus ZCr_b \quad (3.39)$$

kde \oplus značí operaci XOR. Rovnice 3.40 [35] vyjadřuje výpočet této difference pro blok.

$$DZ_b = \frac{1}{8 \times 8} \sum DZC_b \quad (3.40)$$

Následně zprůměrujeme bloky v edge a no-edge oblastech 3.41[35].

$$DZ_{be} = \frac{1}{N_e} \sum_{b=1}^{N_e} DZ_{be}, \quad DZ_n = \frac{1}{N_n} \sum_{i=1}^{N_n} DZ_{bn} \quad (3.41)$$

Nyní vyjádříme disparity feature pro sub-segment 3.42 a temporal segment 3.43 [35].

$$DZ_{es} = \frac{1}{F} \sum_{f=1}^F DZ_{ef}, \quad DZ_{ns} = \frac{1}{F} \sum_{f=1}^F DZ_{nf} \quad (3.42)$$

$$DZ_{et} = \max(DZ_{es}),, s = 1, 2, 3, 4, 5 \quad (3.43)$$

3.4.4 Temporal features

Při výpočtu parametru temporal information (TI) se edge a no-edge oblasti stanoví pouze v prvním snímku sub-segmentu. Temporal feature pro edge oblast sub-segmentu je definován rovnicemi 3.44 a 3.45[35],

$$TI_{es}(m, n, t_k) = \frac{1}{N_e} \sqrt{\frac{1}{8 \times 8} \sum |x_e(m, n, t_f) - x_e(m, n, t_{f+k})|^2} \quad (3.44)$$

$$TI_{es} = \sum_{k=1}^3 TI_{es}(m, n, t_k) \quad (3.45)$$

kde t_f a t_{f+k} jsou první a následující snímky v sub-segmentu, N_e je počet edge bloků a $x_e(m, n)$ je edge oblast. Stejným způsobem vypočítáme TI_{ns} pro no-edge oblast. Následně vypočítáme temporal feature pro temporal segment TI_{et} jako maximum jednotlivých segmentů 3.46[35].

$$TI_{et} = \max(TI_{es}), \quad s = 1, 2, 3, 4, 5 \quad (3.46)$$

Podobně získáme TI_{nt} . Postup 3.44-3.46 uplatníme na levý i pravý snímek, tím získáme RTI_{et} , LTI_{et} , RTI_{nt} a LTI_{nt} . Temporal feature pro edge a no-edge oblasti pro sub-segment vypočítáme podle rovnice 3.47[35].

$$TI_{te} = \frac{LTI_{et} + RTI_{et}}{2}, \quad TI_{tn} = \frac{LTI_{nt} + RTI_{nt}}{2} \quad (3.47)$$

3.4.5 Hodnocení kvality - STS

Nyní máme vyjádřeny parametry a je potřeba zkombinovat je do výsledného hodnocení kvality. To je popsáno v rovnicích 3.48, 3.49, 3.50, 3.51, 3.52 [35].

$$B_t = (B_{et})^{w1} + (B_{nt})^{w2} \quad (3.48)$$

$$Z_t = (Z_{et})^{w3} + (Z_{nt})^{w4} \quad (3.49)$$

$$DZ_t = (DZ_{te})^{w5} + (DZ_{tn})^{w5} \quad (3.50)$$

$$TI_t = (TI_{te})^{w7} + (Z_{tn})^{w8} \quad (3.51)$$

$$S = \alpha \cdot DZ_t + \beta \cdot B_t \cdot Z_t + \gamma \cdot TI_t \quad (3.52)$$

kde $\alpha, \beta, \gamma, w_1, \dots, w_8$ vhodně nastavíme regresní metodou za použití testovacích dat. V tomto případě byla použita metoda Particle swarn optimalization. Hodnoty S jsou následně upraveny převodní funkcí 2.4 [20], aby bylo možné porovnání se subjektivními testy pomocí korelačních koeficientů. Touto upravou získáme objektivní odhad MOS_p. Koeficienty b_1, b_2, b_3 této převodní funkce jsou vhodně nastaveny regresní metodou.

Kapitola 4

Databáze

Při výzkumu, který se zabývá stereoskopickým videem, je zásadní mít k dispozici databázi videí pro ověření hypotéz. Většina databází, které se používají k tomuto účelu, se skládá z videí, jež trvají několik vteřin. Co se týká obsahu, měl by být různorodý. Také je běžné, že každé video je v databázi v několika verzích, které se liší použitou kompresní metodou, přidaným zkráslením, nastavením snímacího aparátu atd. Pro účely této práce byly použity dvě veřejně dostupné databáze, ke kterým byly poskytnuty výsledky subjektivních testů.

4.1 Datáze EPFL

Tato databáze je podrobně popsána v literatuře [38]. Na snímání byly použity dvě identické komerčně dostupné kamery s HD rozlišením, které byly od sebe posunuté v horizontálním směru. U obou kamer bylo použito stejné nastavení. V případě nastavení přiblížení (zoom) bylo tohoto docíleno použitím nastavení v krajních polohách. Dále byla zajištěna synchronizace kamer.

Následně byly sejmuté sekvence upravené algoritmy na potlačení prostorového zkráslení a vyvážení bílé barvy.

Databáze se skládá z 10 scén, z toho 6 je veřejně dostupných. Každá scéna má jiné horizontální posunutí kamer v rozmezí 10-50cm, celkově se tedy jedná o 30 videí. Sekvence mají různá rozlišení, 25fps a 10s.

Subjektivních testů se zúčastnilo 20 osob (6 žen, 14 mužů) ve věku 24-37 s průměrem 27. Schopnost zúčastněných byla prověřena na Snellenově tabuli a testem barvosleposti Ishihara. Test proběhl metodou SS - single stimulus a byla použita stupnice o 5 stupních (excellent, good, fair, poor, bad).

4.2 Databáze NAMA3DS1-COSPAD1

Tato databáze je podrobně popsána v literatuře [39]. Pro snímání byla použita polo-profesionální stereo kamera, která má objektivy ve vzdálenosti 6cm od sebe. Nebylo tedy nutné zavádět stejná omezení jako v případě databáze EPFL.

Databáze obsahuje 11 scén, každá scéna je přítomna v 10 verzích. Zkratka COSPAD značí coding and spatial distortion, tedy prostorové zkreslení a zkreslení kódováním. Jednotlivé verze se liší přítomností těchto zkreslení. Jde o kodeky H.264 a JPEG2K, zkreslení v prostorové doméně je zastoupeno převzorkováním a rozostřením. Jedno video z každé scény je ponecháno bez zkreslení a slouží jako reference.

Subjektivních testů se zúčastnilo 29 osob(12 žen, 17 mužů) ve věku 18-63. Schopnost zúčastněných byla prověřena na Snellenově tabuli, testem barvosleposti Ishihara, testem míry stereopse Randot a testem oční dominance Porta. Test proběhl metodou ACR-HR - Absolute category rating - hidden reference a byla použita stupnice o 5 stupních (excellent, good, fair, poor, bad).

Kapitola 5

Implementace a Výsledky

Metody byly implementovány v prostředí MATLAB. Přehled a krátký popis skriptů použitých při implementaci je v příloze B.

5.1 Using disparity for quality assesment of stereoscopic images - 2008

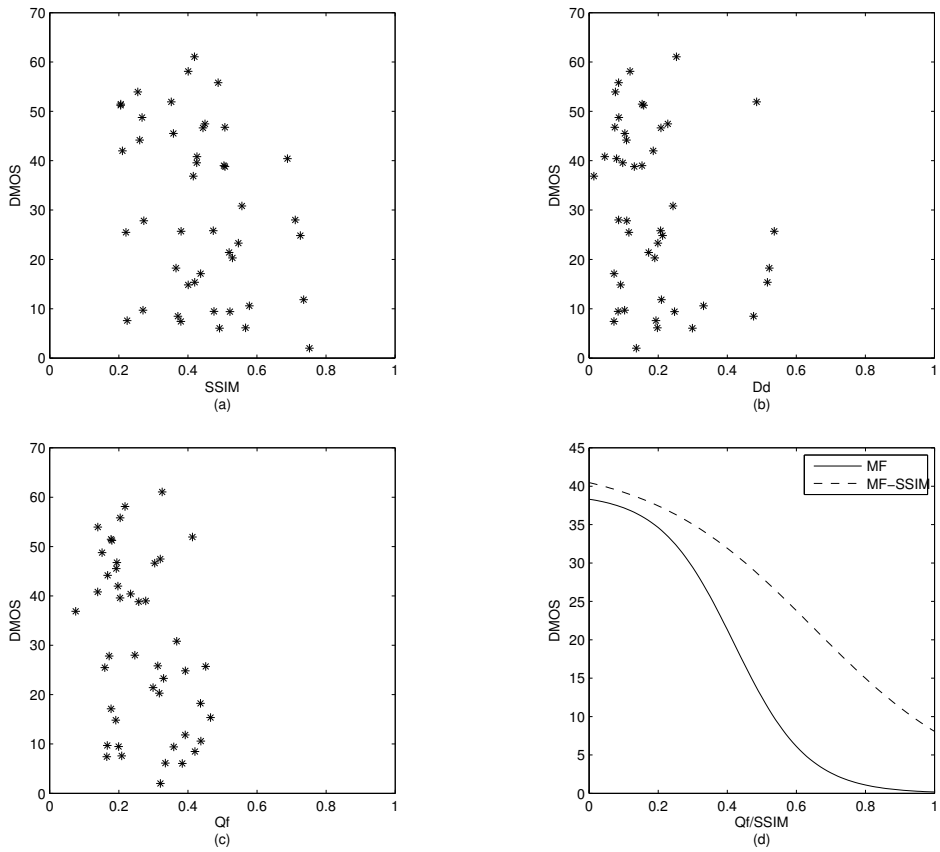
Při implementaci této metody byl použit skript na výpočet SSIM [19] a implementaci algoritmu Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure [40] pro výpočet disparitního obrazu. Na nastavení koeficientů převodní funkce posloužila funkce *nlinfit.m*, která je součástí Matlab Statistics Toolbox.

Protože tato metoda pracuje pouze s každým jednotlivým snímkem a nebere v potaz relaci mezi snímky, postačí na vyhodnocení výsledků databáze statických snímků. Z důvodu snížení výpočetní náročnosti bylo využito skutečnosti, že obě databáze popsané v kapitole 4 mají variantu pro statické snímky.

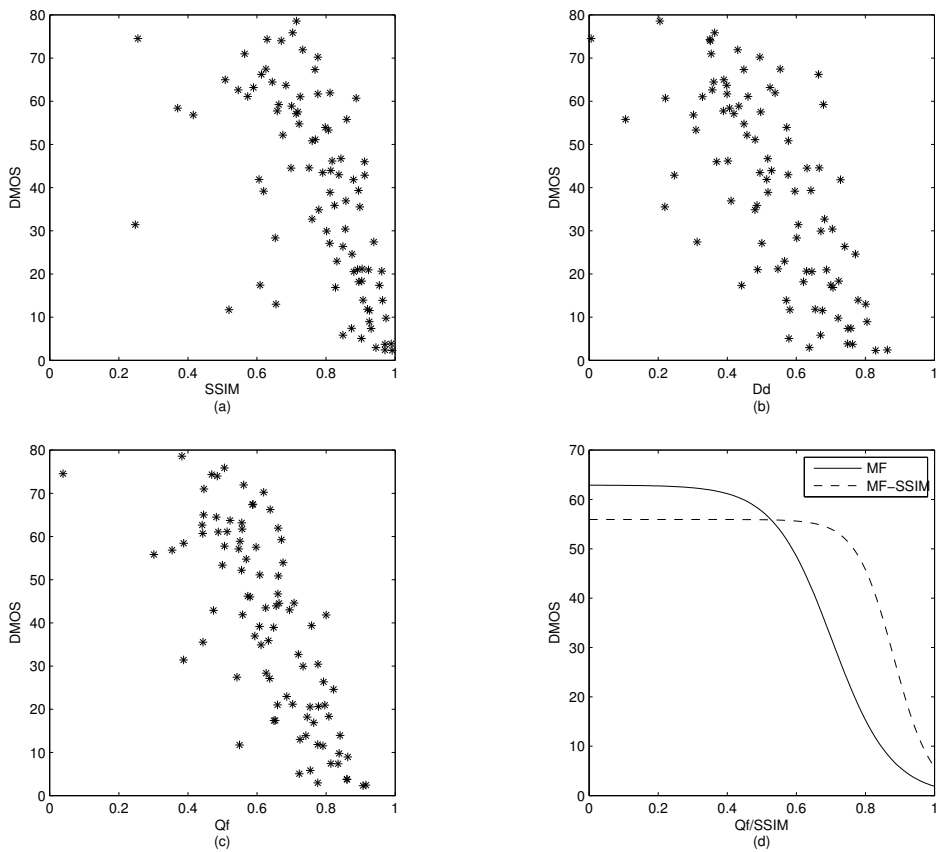
Některé snímky z databáze NAMA3DS1 musely být oříznuty, protože obsahovaly pruhy stejné barvy při okrajích snímku. Bez oříznutí by tyto pruhy zvyšovaly hodnotu SSIM indexu.

Na obr. 5.1.a-c je znázorněna korelace dílčích parametrů metody se subjektivními testy DMOS pro databázi EPFL. Obr 5.1.d znázorňuje optimalizovanou převodní funkci 2.4 [20]. Z těchto grafů je patrné, že pro tuto databázi je korelace se subjektivními testy velmi malá. Je to způsobeno tím, že porovnáváme mezi sebou snímky, které se liší ve vzdálenosti mezi fotoaparáty při snímání. Každá verze snímku zachycuje danou scénu z jiného úhlu. V tomto případě kritéria SSIM a PCC selhávají.

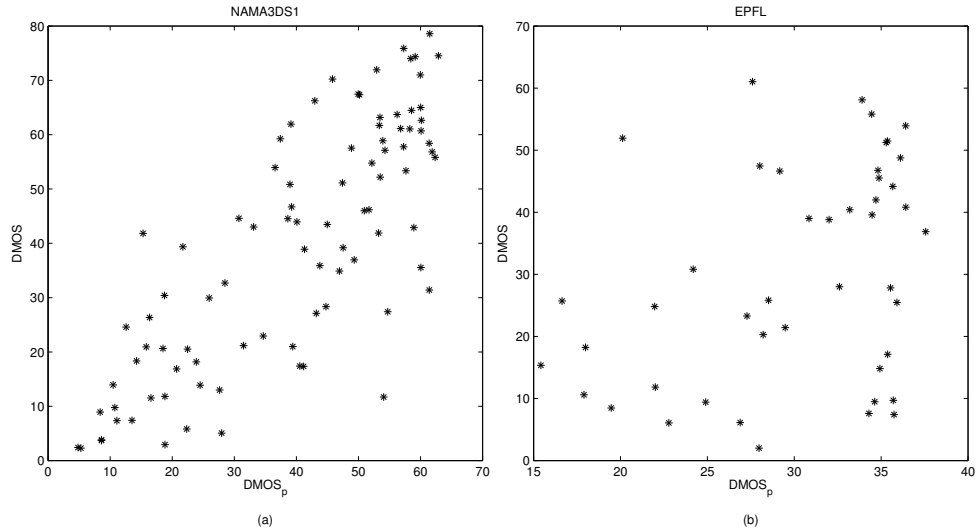
Stejné parametry jsou pro databázi NAMA3DS1 zobrazeny na obr. 5.2. V tomto případě se jednotlivé verze snímku liší použitou kompresí nebo přidaným zkreslením, což umožňuje použití SSIM s větším úspěchem. Na obr. 5.3 je vykreslena korelace mezi výsledky subjektivních testů DMOS a jejich odhadem DMOS_p.



Obrázek 5.1: Znázornění korelace mezi DMOS a dílčími parametry, databáze EPFL



Obrázek 5.2: Znázornění korelace mezi DMOS a dílčími parametry, databáze NAMA3DS1



Obrázek 5.3: DMOS vs DMOS_p, (a) - EPFL, (b) NAMA3DS1

Výsledky této metody jsou shrnuty v tab. 5.1. Pro porovnání jsou zde také uvedeny výsledky metody SSIM.

	NAMA3DS1		EPFL	
	Qf	SSIM	Qf	SSIM
PCC	0.7956	0.7315	0.3669	0.3177
SC	0.7634	0.6878	0.3179	0.3312
OR	0.4778	0.4556	0.5556	0.6222
RMSE	13.3663	16.2006	15.8977	16.2002

Tabulka 5.1: Porovnání odhadnutých výsledků se subjektivními testy

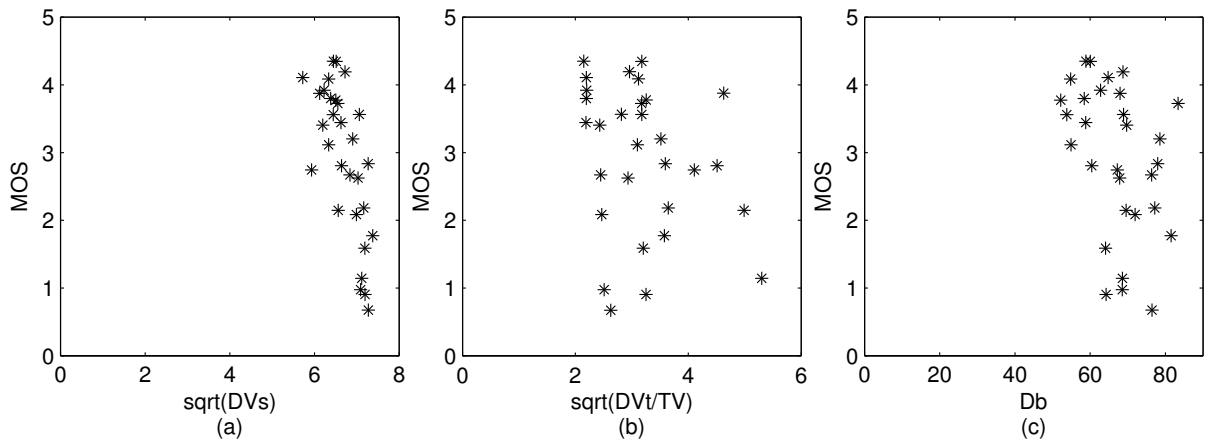
5.2 Stereoscopic video quality perception model - 2011

Při implementaci bylo nutné vypočítat pohybový vektor a disparitní obraz pro každý snímek, přes tyto snímky se následně průměruje. Výpočet těchto parametrů je poměrně náročný, proto jsem tyto parametry vypočítal pro každý desátý snímek.

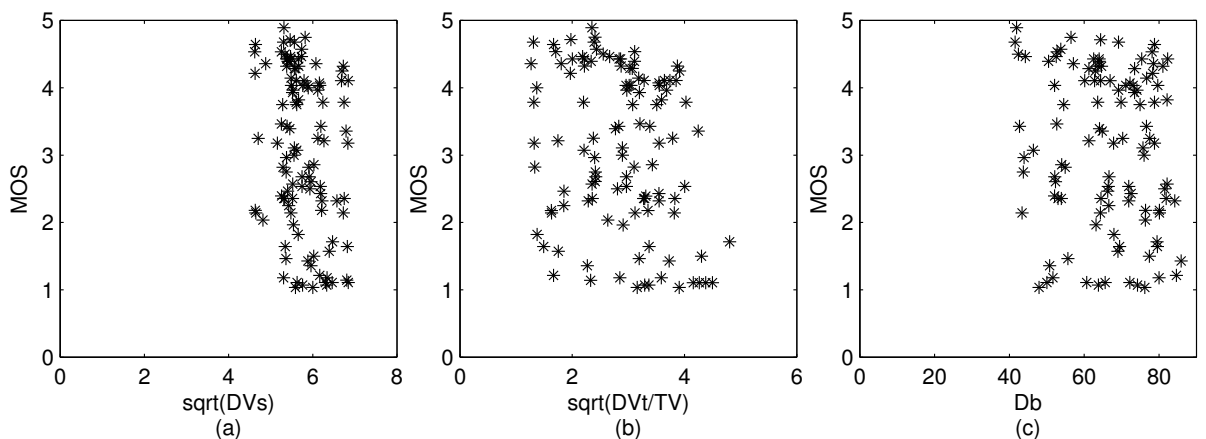
Pohybový vektor byl odhadnut z levých snímků algoritmem popsáním v literatuře [41]. Pro odhad disparity byl použit stejný algoritmus, jako u předchozí metody [40]. Parametry, ze kterých se kombinuje výsledné hodnocení kvality jsou poté extrahovány z disparity a pohybového vektoru. Na lineární regresi byl použita funkce `fitlm.m`, která je součástí Matlab Statistics Toolbox. Jako trénovací databáze bylo náhodně vybráno 60% videí z databáze.

Na obr. 5.4 jsou vyobrazeny korelace dílčích nezávislých parametrů, ze kterých se bude

kombinovat výsledné hodnocení kvality, pro databázi EPFL. Stejné parametry pro databázi NAMA3DS1 - COSPAD1 jsou vyobrazeny na obr. 5.5

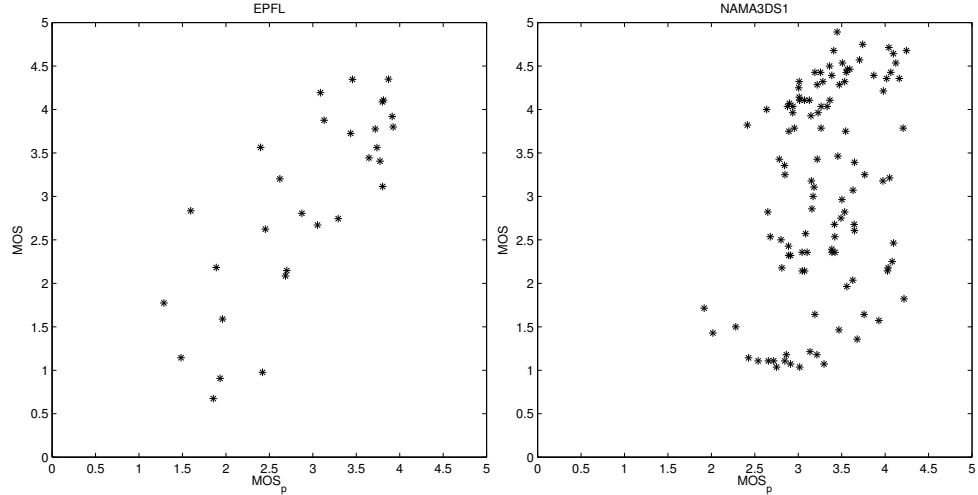


Obrázek 5.4: MOS vs SVQPAM, databáze EPFL



Obrázek 5.5: MOS vs SVQPAM, databáze NAMA3DS1

Z těchto grafů je patrné, že pro databázi EPFL je korelace parametrů, zejména parametru DV_s , vyšší než je tomu u databáze NAMA3DS1. Tyto dílčí parametry jsou extrahovány především z disparitního obrazu. Videá zachycující stejnou scénu v databázi EPFL se liší vzdáleností kamer, tudíž hodnoty disparitního obrazu budou pro tato videa rozdílné. V případě databáze NAMA3DS1, kde se videa zachycující stejnou scénu liší degradací obrazu, se tato degradace přenesou do disparitního obrazu. Parametry vypočítané z tohoto degradovaného disparitního obrazu se subjektivními výsledky korelují pouze málo. Tato metoda se zaměřuje na jevy spojené s dojemem hloubky a pohybem. Není ale navržena tak, aby byla schopna postihnout degradace v obraze. Na obr. 5.6 je vykreslena korelace mezi výsledky subjektivních testů MOS a jejich odhadem MOS_p pro obě databáze. Výsledky této metody jsou shrnuty v tab. 5.2.



Obrázek 5.6: MOS vs MOS_p, (a) - EPFL, (b) NAMA3DS1

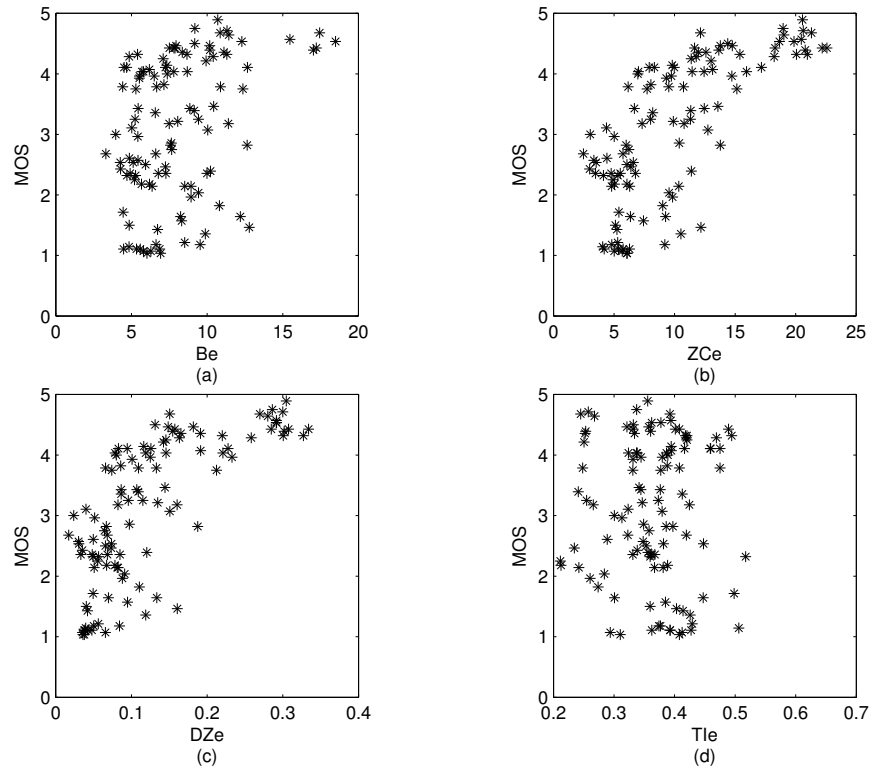
	EPFL			NAMA3DS1		
	test	train	all	test	train	all
PCC	0.7600	0.8084	0.7902	0.2507	0.4105	0.3373
SC	0.6993	0.7069	0.7673	0.2407	0.4359	0.3390
OR			0,2333			0.8182
RMSE			0.6647			1.1286

Tabulka 5.2: Porovnání odhadnutých výsledků se subjektivními testy - SVQPM

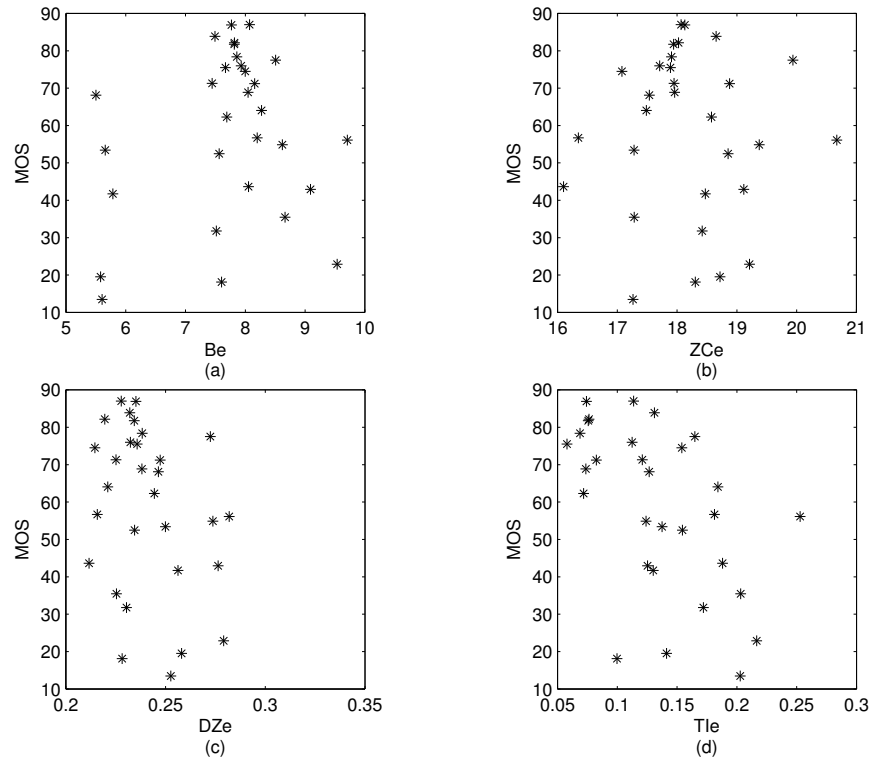
5.3 Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction 2010

Tato metoda počítá s průběžným hodnocením kvality, kdy se vyhodnocuje kvalita pro temporal segment, který zahrnuje 20 snímků. Parametr pro temporal segment je vyjádřen jako aritmetický průměr parametrů těchto snímků (B, Z, 3.38) nebo jako kombinace průměru a maxima (TI 3.43, DZ 3.46). Protože máme k dispozici databázi pouze se subjektivním hodcením pro celou video sekvenci, budeme uvažovat parametry pro celou sekvenci jako aritmetický průměr temporal segmentů pro tuto sekvenci.

Dílčí parametry této metody jsou vypočítány pomocí funkce *STS_epfl.m* pro databázi EPFL a *STS_NAMA3DS1.m* pro databázi NAMA3DS1. Jejich korelace se subjektivními testy je vyobrazena na obr. 5.7 pro edge oblasti v databázi NAMA3DS1 a na obr. 5.8 pro no-edge oblasti v databázi EPFL.



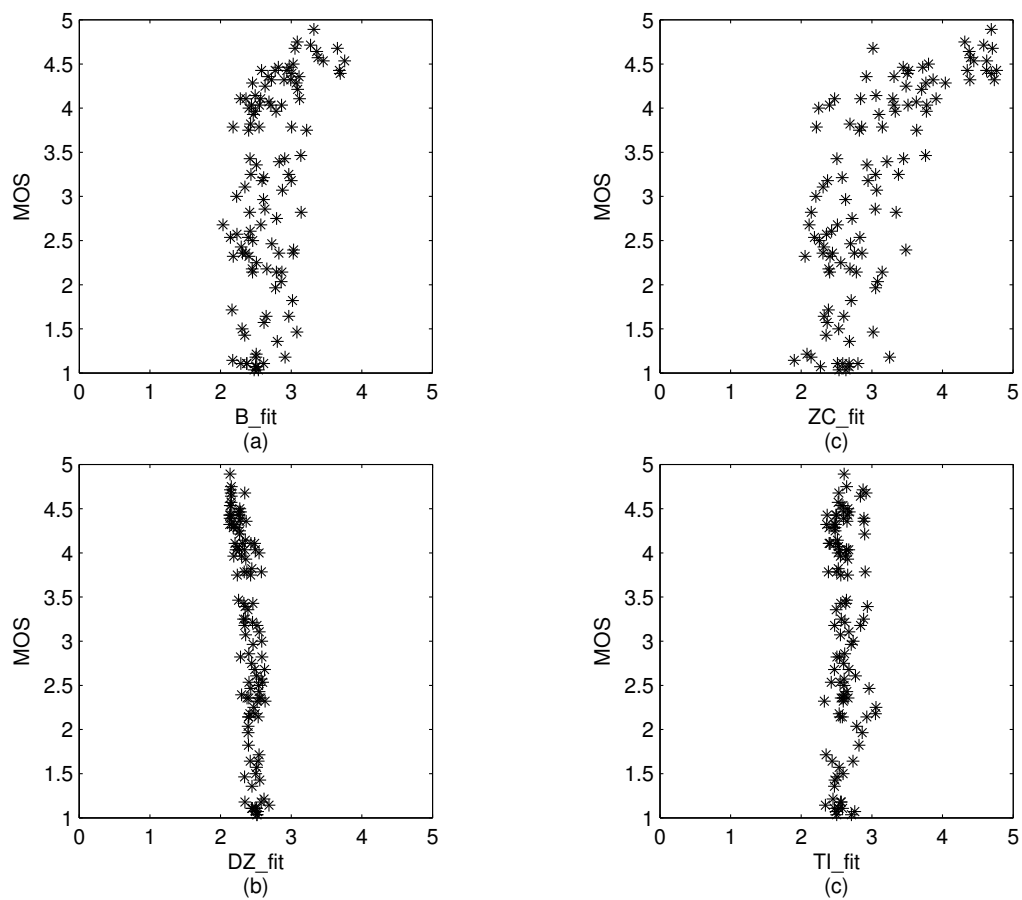
Obrázek 5.7: Znázornění korelace mezi MOS a dílčími parametry metody, edge area, databáze NAMA3DS1



Obrázek 5.8: Znázornění korelace mezi MOS a dílčími parametry metody, edge area, databáze EPFL

Na obr 5.8 je vidět, že pro databázi EPFL dílčí parametry nekorelují. V tomto případě je tato korelace tak malá, že není možné provést kombinaci těchto parametrů do celkového hodnocení kvality.

Kombinace dílčích parametrů se do celkového hodnocení kvality pomocí Particle Swarm algoritmu. Protože u tohoto algoritmu bylo velice obtížné určit predikci jeho koeficientů, byla použita funkce nelineární regrese *nlinfit.m*, která je součástí Matlab Statistics Toolbox. Nastavení koeficientů bylo provedeno ve dvou krocích. První krok byl odhad koeficientů w_1, \dots, w_8 v rovnicích 3.48, 3.49, 3.50, 3.51. K tomu posloužila funkce *X22.m* a jako testovací data bylo náhodně vybráno 60% videí z databáze. Na obr. 5.9 je vyobrazena korelace takto upravených parametrů se subjektivními testy.



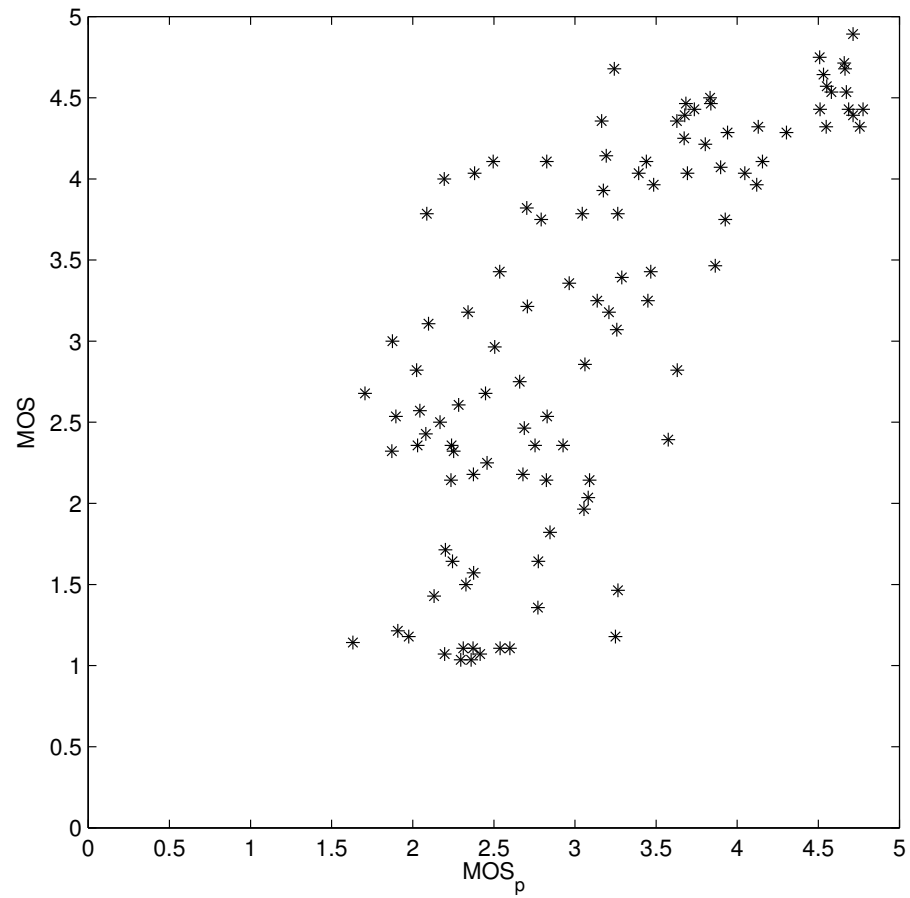
Obrázek 5.9: Znáornění korelace mezi DMOS a dílčími parametry metody, databáze NAMA3DS1

Poté proběhl odhad koeficientů α, β, γ z rovnice 3.51 [35] optimalizací funkce *S_fit.m*, přičemž byla použita stejná testovací data. Regresní koeficienty jsou shrnuty v tab.5.3.

w1	w2	w3	w4	w5	w6	w7	w8	α	β	γ
0.2605	0.5700	0.1652	0.3471	0.1045	0.0609	-0.5172	-0.0735	-1.1469	0.1330	1.5996

Tabulka 5.3: Regresní koeficienty

Na obr. 5.10 je vykrelena korelace odhadnutých MOS_p se subjektivními MOS pro databázi NAMA3DS1. Výsledky metody jsou shrnuty v tab. 5.4.



Obrázek 5.10: MOS vs MOS_p , (a) - EPFL, (b) NAMA3DS1

OR	0.5727
RMSE	0.8192
PCC	0.7136
<i>PCC_test</i>	0.6954
<i>PCC_train</i>	0.7120
<i>SC</i>	0.7149
<i>SC_test</i>	0.6973
<i>SC_train</i>	0.6572

Tabulka 5.4: Výsledky metody

Kapitola 6

Závěr

Tato diplomová práce zpracovává přehled metod pro hodnocení kvality stereoskopického videa. Metody byly vybrány, tak aby tento přehled byl uceleným shrnutím současného stavu dané problematiky. Byla tedy snaha vybrat z každé kategorie metod její nejlepší zástupce.

Následně byly vybrané metody implementovány v prostředí Matlab. Účinnost těchto metod byla ověřena na dvou veřejně dostupných databázích. Při implementacích se bohužel nepodařilo dosáhnout stejně dobrých výsledků jako v literatuře.

Zajímavé jsou ale výsledky týkající se vhodnosti metody pro danou databázi. Databáze použité na ověření účinnosti měly totiž úplně rozdílné vlastnosti. V databázi EPFL je hlavním parametrem, který určuje kvalitu dané sekvence vzdálenost kamer při snímání. Z výsledků shrnutých v tab. , a je patrné, že pro tuto databázi dosahuje nejlepších výsledků metoda SVQPM, která je popsána v kapitole 3.1. Tato metoda extrahuje své parametry právě z informace o disparity a proto má pro tuto databázi dobré výsledky.

Naopak pro druhou použitou databázi NAMA3DS1, která je založena na prostorovém zkreslení a zkreslení kódováním tato metoda selhává. Zbylé dvě metody popsané v kap. a které své dílčí parametry extrahují z obrazu jsou schopny odhadnout hodnocení kvality pro tuto databázi lépe.

Je tedy patrné, že dnes známé metody jsou poměrně úzce specializované a většinou pokrývají pouze některé z aspektů, které jsou pro stereoskopické video běžné. Tato specializace souvisí s tím, že tato problematika je poměrně komplexní a výpočetně náročná. Ani tyto specializované metody bohužel nedosahují uspokojivých výsledků.

Literatura

- [1] PIETER J.H. SEUNTIËNS Visual Experience of 3D TV. Eindhoven, 2006. ISBN 90-386-0705-9. Phd thesis. Technische Universiteit Eindhoven. Vedoucí práce prof.dr.ir. C.J. van Duijn.
- [2] CHARLES WHEATSTONE Contributions to the Physiology of Vision, Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision, In *Phil. Trans. R. Soc. Lond.* 1838 128, doi: 10.1098, rstl.1838.0019, 1838
- [3] SMOLIC, A.; MUELLER, K.; MERKLE, P.; VETRO, A.; Development of a new MPEG standard for advanced 3D video applications, In *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on* , vol., no., pp.400-407, 16-18 Sept. 2009
- [4] HANHART, P.; DE SIMONE, F.; EBRAHIMI, T.; Quality assessment of asymmetric stereo pair formed from decoded and synthesized views, In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on* , vol., no., pp.236-241, 5-7 July 2012
- [5] JIN, L.; BOEV, A.; GOTCHEV, A.; EGIAZARIAN, K.; 3D-DCT based perceptual quality assessment of stereo video, In *Image Processing (ICIP), 2011 18th IEEE International Conference on* ,. vol., no., pp.2521-2524, 11-14 Sept. 2011
- [6] MITTAL, A.; MOORTHY, A.K.; GHOSH, J.; BOVIK, A.C.; , Algorithmic assessment of 3D quality of experience for images and videos, In *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE* vol., no., pp.338-343, 4-7 Jan. 2011
- [7] JONG-SEOK LEE; DE SIMONE, F.; EBRAHIMI, T.; RAMZAN, N.; IZQUIERDO, E. Quality assessment of multidimensional video scalability, In *Communications Magazine, IEEE*, vol.50, no.4, pp.38-46, April 2012
- [8] A. BOEV, D. HOLLOSI, A. GOTCHEV. Classification of stereoscopic artefacts. Technical report D5.1. July 2008

- [9] ANDREW WOODS, TOM DOCHERTY, ROLF KOCH. Image Distortions in Stereoscopic Video Systems. In *Proceedings of the SPIE Volume 1915, Stereoscopic Displays and Applications IV*. San Jose, California. February 1993.
- [10] FRANTIŠEK MIKŠÍČEK Causes of Visual Fatigue and Its Improvements in Stereoscopy. Technical Report No. DCSE/TR-2006-04. University of West Bohemia in Pilsen. Pilsen, 2006.
- [11] DOUTRE, C.; NASIOPOULOS, P.; Optimized contrast reduction for crosstalk cancellation in 3D displays, In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011* ,. vol., no., pp.1-4, 16-18 May 2011
- [12] BARTCZAK, B.; VANDEWALLE, P.; GRAU, O.; BRIAND, G.; FOURNIER, J.; KERBIRIOU P.; MURDOCH, M.; MULLER, M.; GORIS, R.; KOCH, R.; VAN DER VLEUTEN, R.; Independent 3D-TV Production and Delivery Using the Layered Depth Video Format, In *Broadcasting, IEEE Transactions on* , vol.57, no.2, pp.477-490, June 2011
- [13] SOLH, M.; ALREGIB, G.; , A no-reference quality measure for DIBR-based 3D videos, In *Multimedia and Expo (ICME), 2011 IEEE International Conference on* , vol., no., pp.1-6, 11-15 July 2011
- [14] BOSCH, E.; LE CALLET, P.; MORIN, L.; PRESSIGOUT, M. An edge-based structural distortion indicator for the quality assessment of 3D synthesized views, In *Picture Coding Symposium (PCS)*, vol., no., pp.249-252, 7-9 May 2012
- [15] HIRSCHMÜLLER H.; SCHARSTEIN D.; Evaluation of cost functions for stereo matching, In *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, Minneapolis, MN, June 2007
- [16] ZITNICK L. C.; KANG S. B.; UYTENDAELE M.; WINDER S.; SZELISKI R.; High-quality video view interpolation using a layered representation, Interactive Visual Media Group, Microsoft Research, Redmond, WA
- [17] TANGER, R.; 3D4YOU [online], Berlin, 2009. Dostupné z: < [http : //tech.ebu.ch/docs/events/3dtv09/presentations/ebu_3dtv09_tanger.pdf](http://tech.ebu.ch/docs/events/3dtv09/presentations/ebu_3dtv09_tanger.pdf) >
- [18] MASAYUKI TANIMOTO; FTV (Free viewpoint TV) and Creation of Ray-Based Image Engineering, In *ECTI Transaction on Electrical Engineering, Electronics and Communications*, Vol. 6, No. 1, pp.3-14, February 2008.
- [19] ZHOU WANG; LIGANG LU; BOVIK, A.C.; Video quality assessment using structural distortion measurement, In *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol.3, no., pp. III-65- III-68 vol.3, 2002

- [20] BENOIT, A.; LE CALLET, P.; CAMPISI, P.; COUSSEAU, R.; PRESSIGOUT, M.; MORIN, L. Using disparity for quality assessment of stereoscopic images, In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* , vol., no., pp.389-392, 12-15 Oct. 2008
- [21] FELZENSZWALB, P.F.; HUTTENLOCHER, D.R.; Efficient belief propagation for early vision, In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* , vol.1, no., pp. I-261- I-268 Vol.1, 27 June-2 July 2004
- [22] J. L. RODGERS AND W. A. NICEWANDER.; Thirteen ways to look at the correlation coefficient, In *The American Statistician*,. 42(1):59–66, February 1988
- [23] SARIKAN, S.S.; OLGUN, R.F.; AKAR, G.B.; Quality evaluation of stereoscopic videos using depth map segmentation, In *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on* , vol., no., pp.67-71, 7-9 Sept. 2011
- [24] DELON, J.; DESOLNEUX, A.; LISANI, J.-L.; PETRO, A.B.; A Nonparametric Approach for Histogram Segmentation, In *Image procesing IEEE Transaction on* , . vol.16, no.1, pp.253-261 Jan. 2007
- [25] O. STANKIEWICZ; K. WEGNER, An enhancement of Depth Estimation Reference Software with use of soft-segmentation, In *International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11* Scottsdale, USA, 2006
- [26] DONGHYUN KIM; SEUNGCHUL RYU; KWANGHOON SOHN; Depth perception and motion cue based 3D video quality assessment, In *Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium on* , . vol., no., pp.1-4, 27-29 June 2012
- [27] KWANGSUNG H.; MUNCHURL K. A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video. In *Image Processing (ICIP), 2011 18th IEEE International Conference on* , . vol., no., pp.2525-2528 11-14 Sept. 2011
- [28] LIYUAN XING; JUNYONG YOU; EBRAHIMI, T.; PERKIS, A.; Assessment of Stereoscopic Crosstalk Perception, In *Multimedia, IEEE Transactions on* , vol.14, no.2, pp.326-337, April 2012
- [29] KLAUS, A.; SORMANN, M.; KARNER, K.; , Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure, In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* , vol.3, no., pp.15-18, 2006

- [30] YAO NIE; KAI-KUANG MA; , Adaptive rood pattern search for fast block-matching motion estimation, In *Image Processing, IEEE Transactions on*, vol.11, no.12, pp. 1442- 1449, Dec 2002
- [31] K. EGIAZARIAN, J. ASTOLA, N. PONOMARENKO, V. LUKIN, F. BATTISTI, AND M. CARLI, New full-reference quality metrics based on HVS, In *Int. Workshop on Video Proc. and Quality Metrics*, Scottsdale, USA, 2006
- [32] <http://www.r-project.org/>
- [33] SAZZAD, Z.M.P.; YAMANAKA, S.; HORITA, Y. Continuous stereoscopic video quality evaluation In *Proc. SPIE*,. vol 7524, no., pp.106-111 18-21 Jan 2010, San Jose, USA
- [34] BOSCH, E.; PEPION, R.; LE CALLET, P.; KOPPEL, M.; NDJIKI- NYA, P.; PRESSIGOUT, M.; MORIN, L. Towards a New Quality Metric for 3-D Synthesized View Assessment, In *Selected Topics in Signal Processing, IEEE Journal of*, vol.5, no.7, pp.1332-1343, Nov. 2011
- [35] SAZZAD, Z.M.P.; YAMANAKA, S.; HORITA, Y. Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on* , vol., no., pp.106,111, 21-23 June 2010
- [36] R. AKHTER, Z. M. PARVEZ SAZZAD, Y. HORITA, AND J.BALTES "No-reference stereoscopic image quality assessment," In *Proc. SPIE*, vol. 7524, Jan. 18-21, San Jose, USA, 2010.
- [37] R. AKHTER Perceived Image Quality Assessment for Stereoscopic Vision Winnipeg, Manitoba, Canada, May 2011, Dostupné z: <http://aalab.cs.umanitoba.ca/andersj/Publications/pdf/RoushainAkhterMSc.pdf>, Master Thesis, Department of Computer Science The University of Manitoba, Vedoucí práce Dr. Jacky Baltes.
- [38] LUTZ GOLDMANN, FRANCESCA DE SIMONE, TOURADJ EBRAHIMI, Impact of Acquisition Distortions on the Quality of Stereoscopic Images In *5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)* Scottsdale, USA, 2010.
- [39] MATTHIEU URVOY, JESÚS GUTIÉRREZ, MARCUS BARKOWSKY, ROMAINCOUSSEAU, YAO KUDOTA, VINCENT RICORDEL, PATRICK LE CALLET AND NARCISO GARCÍA, NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences In *Fourth International on Quality of Multimedia Experience*, Yarra Valley, July 2012.

- [40] LANKTON, SHAWN. 3D Vision with Stereo Disparity. [online]. [cit. 2013-12-12] Dostupné z: [http : //www.shawnlankton.com/2007/12/3d - vision - with - stereo - disparity/](http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/)
- [41] CHAN, S.H.; VO, D.T.; NGUYEN, T.Q., Subpixel motion estimation without interpolation, In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* , [online]. [cit. 2013-11-11]. Dostupné z: [http : //www.mathworks.com/matlabcentral/fileexchange/44244 - subpixel - motion - estimation - without - interpolation/content/Motion_{Est}.m](http://www.mathworks.com/matlabcentral/fileexchange/44244-subpixel-motion-estimation-without-interpolation/content/MotionEst.m) vol., no., pp.722,725, 14-19 March 2010

Příloha A

Obsah přiloženého CD

Na přiloženém cd se nachází následující dokumenty

- Adresář literatura: citovaná literatura v elektronické formě
- Adresář latex: zdrojový kód z latexu
- Adresář dip: elektronická verze této práce v pdf.
- Adresář matlab: zdrojové kódy skriptů v jazyce MATLAB

Příloha B

Přehled použitých skriptů v jazyce MATLAB

Začneme přehledem skriptů, které byly převzaty.

- *ssim_index.m* [19] výpočet metody hodnocení kvality s referencí SSIM - Structural similarity index.
- *stereo.m* [40] výpočet hodnoty disparity.
- *Motion_Est.m* [41] výpočet pohybového vektoru
- *stereo_database.m* - Tento skript vypočte hodnotu disparity pro celou databázi a výsledek uloží na disk ve formátu MAT. Pro vlastní výpočet disparity je zde použit algoritmus z literatury [19].
- *motionVect_Database.m* - Tento skript vypočte hodnotu pohybového vektoru pro celou databázi a výsledek uloží na disk ve formátu MAT. Pro vlastní výpočet pohybového vektoru je zde použit algoritmus z literatury [41]
- *SVQPM.m* - Tento skript vypočte dílčí parametry metody popsané v kapitole 6
- *STS_vysledky_NAMA3DS1.m* - Tento skript vypočte výsledné hodnocení kvality z dílčích parametrů pro databázi NAMA3DS1.
- *STS_vysledky_epfl.m* - výpočet výsledného hodnocení kvality z dílčích parametrů pro databázi epfl.
- *disparity_im.m* - výpočet disparity pro metodu Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction
- *mapping.m* - převodní funkce 2.4

- $X_{22.m}$ - regresní funkce 3.51
- S_{fit} - regresní funkce 3.52