

VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice PhD Thesis, sv. 405

ISSN 1213-4198

thesis IS

Ing. Kamil Říha

**Akvizice 3D scény
pro její vyobrazení
na auto-stereoskopickém displeji**

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

ÚSTAV TELEKOMUNIKACÍ

Ing. Kamil Říha

**AKVIZICE 3D SCÉNY
PRO JEJÍ VYOBRAZENÍ NA
AUTO-STEREOSKOPIKÉM DISPLEJI**

**3D SCENE ACQUISITION FOR
AUTO-STEREOSCOPIC DISPLAY**

ZKRÁCENÁ VERZE PH.D. THESIS

Obor: Teleinformatika
Školitel: Prof. Ing. Kamil Vrba, CSc.
Oponenti: Doc. Ing. Karel Bartušek, DrSc.
Doc. Ing. Boris Šimák, CSc.
Datum obhajoby: 20. 2. 2007

Klíčová slova

Epipolární geometrie, DEPTH matice, disparita, dynamické programování, kamera, autostereoskopický, paralaxa, 3D displej, složený obraz, prostorové zobrazení.

Key words

Epipolar Geometry, DEPTH Matrix, Disparity, Dynamic Programming, Camera, Autostereoscopic, Parallax, 3D Display, Combined Picture, Spatial Displaying.

Práce je k dispozici na Vědeckém oddělení děkanátu FEKT VUT v Brně, Údolní 53, Brno, 602 00.

OBSAH

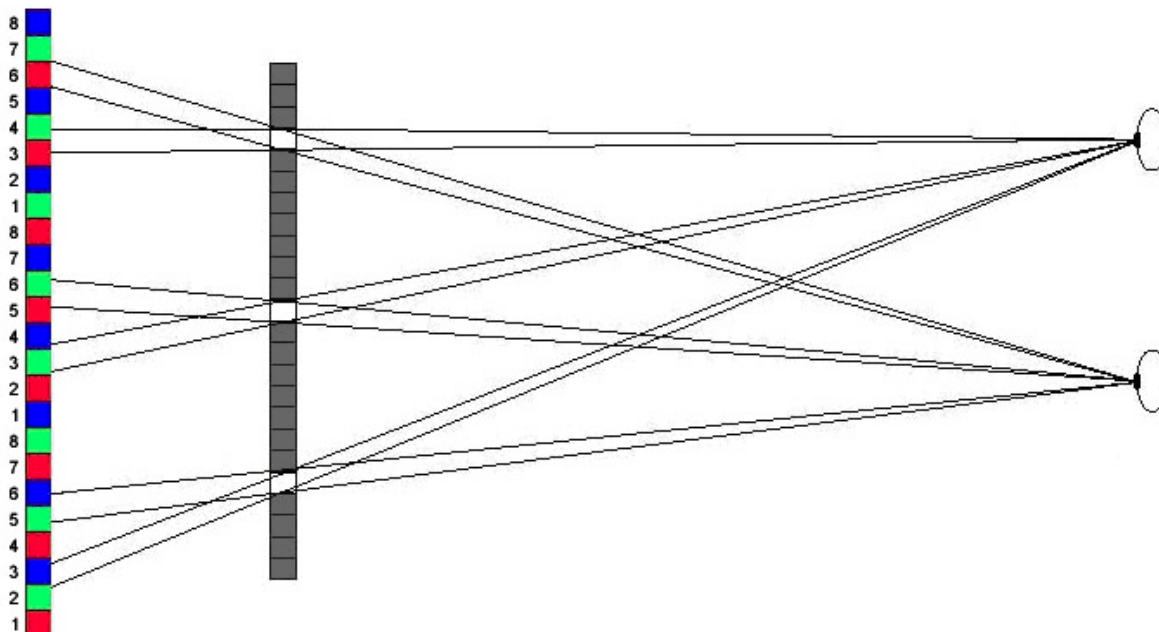
1 ÚVOD.....	5
2 PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY	7
3 CÍLE DISERTAČNÍ PRÁCE	7
4 MINIMÁLNÍ VZDÁLENOST POZOROVATELE OD DISPLEJE	9
5 EFEKTIVNÍ METODA GENEROVÁNÍ NOVÉHO POHLEDU	11
5.1 Jednoduchá stereo geometrie	11
5.2 Určení disparity	12
5.2.1 <i>DSI (Disparity Space Image – disparitní prostor)</i>	12
5.2.2 <i>Dynamické programování</i>	13
5.2.3 <i>Experimentální výsledky</i>	17
5.3 Generování dalších pohledů.....	18
6 ZÁVĚR.....	20
LITERATURA	22
CURRICULUM VITAE	23

1 ÚVOD

Člověk dekóduje prostorovou obrazovou informací z mnoha faktorů mezi nimiž nezanedbatelnou roli hraje i dvourozměrné zpracování různých informací ze scény na základě zkušeností jako jsou například paralaxa, relativní velikost předmětů, zákryty a podobně. Drtivá většina přístupů k řešení problému jak navodit pozorovateli prostorový vjem však využívá stereoskopického principu lidského vnímání prostoru. Cílem je doručit každému oku lišící se obrazy 3D scény, jejichž odlišnost (takzvaná binokulární disparita) spočívá ve změně pozorovacího bodu daného vzdáleností mezi očima – oční bází. Prostředky, kterých bývá k tomuto účelu použito, mohou mít různou podobu. Nejznámější jsou asi stereoskopické brýle, jejichž funkce spočívá ve filtraci příslušného pohledu buď pomocí barvy, anebo polarizace světla. Tyto přístupy však mají jeden základní nedostatek: k dosažení prostorového vjemu je zapotřebí speciálních pomůcek, které mnohdy snižují kvalitu některých parametrů obrazu a musí jimi být opatřeni všichni pozorovatelé. Aby byly odstraněny tyto nevýhody, byly vyvinuty autostereoskopické displeje, které mohou využívat dva a více pohledů, vytvářející před monitorem zóny, ve kterých je prostorového vjemu možné dosáhnout bez jakýchkoli pomůcek. S postupem času se objevilo poměrně mnoho technologií, jak před obrazovkou vytvořit zóny, kde bude každé oko vidět příslušný obraz ze stereo páru. Pravděpodobnost, že se bude oko nacházet v jemu příslušné zóně, však není u dvoupohledových technik příliš vysoká.

Tato práce si bere za cíl zejména předzpracování zdrojových obrazových dat pro kvalitní více pohledové autostereoskopické displeje. Takový displej může pracovat na principu Wavelength-Selective Filter Array (barevně selektivní filtrační pole) a dal by se pak stručně charakterizovat jako osmipohledový bariérový subpixelový autostereoskopický LCD monitor. Osm zdrojových pohledů, které tato jednotka využívá, je kombinováno do složeného obrazu na základě přesně určeného algoritmu uspořádání subpixelů (barevných elementů). Mezi pozorovatele a LCD panel zobrazující speciální složený obraz je umístěn zmíněný WLSFA, který umožňuje na základě relativně jednoduchého fyzikálního principu navození 3D vjemu u pozorovatele.

Na obr. 1 je znázorněn princip, který umožňuje rozdílné adresování pohledů zvláště pro každé oko. Barevný sloupec v levé části obrázku schematicky znázorňuje uspořádání barevných elementů RGB (R-červená, G-zelená, B-modrá) v jednom řádku LCD panelu. Pomocí otvorů v přiloženém filtru dochází k selektivnímu doručení informace z příslušného pohledu na danou scénu, což je symbolicky označeno čísly 1-8 u příslušného subpixelu pro osm zdrojových pohledů.



Obr. 1: Horizontální řez pohledy ze dvou bodů-očí.

Rozbor problematiky geometrie tohoto principu autostereoskopického zobrazování byla autorem této disertační práce publikována v [17]. Ve 4. kapitole této práce je pak podrobně popsána originální publikovaná definice minimální vzdálenosti pro dosažení prostorového vjemu a z ní vyplývající definice zón pro dosažení neznehodnoceného prostorového vjemu.

Pro zobrazení 3D scény tedy existuje poměrně snadno vyrobitelný hardware s minimem nových technologických prvků. Jediným novým prvkem, který je nutné vyrobit, je WLSFA maska, která reflektuje parametry klasického LCD panelu. Problematickou částí této technologie však je oblast akvizice reálných obrazových dat, kdy je pro nasnímání scény obvykle zapotřebí několika kamer umístěných v přesně definovaných pozicích. Zde je široké pole pro další vývoj. Ve 3D počítačové grafice totiž není problém potřebný počet pohledů vytvořit (renderovat), protože daná syntetická scéna je uložena ve formě kompletního prostorového modelu. Snahou ovšem je snímání a prostorové zobrazení skutečné scény. Dokladem toho jsou například 3D biografy, systémy virtuální reality či lékařství. V odborných kruzích jsou shromažďovány podklady pro koncepci 3D televizního řetězce, tento krok je přirovnáván k přechodu od němého k ozvučenému filmu. Ve video praxi se již lze celkem běžně setkat s natáčením pomocí stereo páru kamer, zatímco vývoj akviziční soustavy, která by na výstupu dávala přímo osmici potřebných pohledů pro zobrazovací jednotku, je pro široké využití téměř nemyslitelný. Je zde tedy nutnost vytvoření metody, která handicap více než dvoupohledových autostereoskopických displejů odstraní.

2 PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY

Všech (obecně) n pohledů nutných pro vytvoření složeného prostorového obrazu je možné získávat těmito způsoby:

- Přímým snímáním n pohledů n kamerami. Nevýhodou je obtížná mechanická konfigurace a synchronizované zpracování velkého datového toku z n -tice kamer. Výhodou je, že snímaná obrazová informace není zkreslená aproximací.
- Speciálním tzv. plug-in-em (zásuvným modulem), který potřebné pohledy vytváří (renderuje) z prostorového modelu scény. Jedná se o zpracování prostorového modelu, který je nutné buď synteticky vytvořit, nebo náročným postupem naskenovat 3D objekty v dané scéně. Tato technologie je sice vyvinuta (jako součást konceptu virtuální reality), ale vyznačuje se pro naše účely velkou časovou náročností a redundancí ukládaných dat, nicméně s velkou věrností a zachováním kompletní prostorové informace. Pro zpracování videosekvencí v reálném čase je však takovýto postup prakticky nerealizovatelný.
- Doplněním digitálního 2D snímku uměle vytvořenou hloubkovou (DEPTH) maticí obsahující informaci o hloubce scény pro každý pixel. Umělé vytváření DEPTH matice je většinou řešeno manuálním nebo poloautomatickým definováním hranic oblastí s rozdílnou hloubkou. Při následném generování příslušných pohledů je u všech metod pracujících s DEPTH maticí a jedním obrazem nutné provádět aproximaci v pozicích vzniklých okluzemi (překrytím), což vede ke zkreslení generovaných snímků, ale jako zdroj obrazové informace postačí pouze jeden RGB obraz doplněný hloubkovou maticí ve stupních šedi (většinou s 8-bitovým rozlišením). Principiálně je takový postup vhodný spíše pro statické obrazy.
- Speciální kamerou opatřenou aktivním senzorem pro snímání hloubkové informace například na principu měření doby letu světelného paprsku. Další postup pro vytvoření potřebných pohledů je shodný s předchozím bodem. DEPTH matice je tedy snímána okamžitě s RGB obrazem a zpracování je tak vhodné pro real-timeové zobrazování a přenos jak je zpracováno ve [1]. Obecně však platí, že technika generující nové pohledy z DEPTH matice bude vždy trpět nedostatkem, který spočívá v nutnosti zaplnění děr vzniklých odkrytím okludovaných (zakrytých) částí v referenčním pohledu. Tuto nevýhodu odstraňuje metoda prezentovaná v této práci.

3 CÍLE DISERTAČNÍ PRÁCE

Kromě podrobného rozboru problematiky vícehledového autostereoskopického zobrazování, jehož původní výsledky jsou uvedeny v kapitole 4, bylo hlavní snahou vytvoření původního postupu, který by odstranil obecné nevýhody vícehledového snímání obrazu a s ním souvisejících technik generování nových pohledů popsaných

v kapitole 2. Pro masově využitelné potřeby prostorového zobrazování byla tedy hledána metoda akvizice 3D scény, která:

- nebude potřebovat příliš specializovaný hardware,
- bude schopna v potřebném rozmezí efektivně produkovat pohledy na 3D scénu pro autostereoskopický displej.

V oblasti počítačového vidění je řešena úloha, kdy je prostorová informace extrahována na základě korespondence obrazových bodů a z ní plynoucí binokulární disparity ve dvojici snímků zobrazované scény pořízených s odlišnou perspektivou. Tato dvojice snímků je nazývána stereo pár. Použitím stereo páru kamer jako jednotky pro akvizici 3D scény je splněn první požadavek na jednoduchost hardware. Zbývá tedy nalézt metodu, která bude co nejefektivněji schopna generovat nové pohledy bez explicitního vytváření prostorového modelu.

Vzhledem k současnému stavu problematiky se nabízí možnost zaměřit se primárně na získávání DEPTH matice ze stereo páru a až poté na generování nových pohledů na základě obrazové informace obsažené ve snímku kompatibilním s nalezenou hloubkovou maticí. Nicméně pouhé nalezení DEPTH matice ve formě obvyklé pro počítačové vidění, byť by byla velmi přesná, ještě samo o sobě pro kompletní využití obrazové informace obsažené ve stereo páru nestačí. Jsou zde totiž (v běžných scénách) relativně hojné oblasti částečně okludovaných pixelů zobrazených vždy pouze v jednom obraze ze zdrojové dvojice. Tento fakt tedy spíše vybízí k přímému využití této dvojice bez nutnosti použití aproximace vzniknuvších děr jako je tomu v případě postupů na základě klasické DEPTH matice a jednoho referenčního obrazu. V kapitole 5 je realizace takového efektivního postupu uvedena.

Cíle řešení celého problému by se daly shrnout do následujících bodů:

- **Podrobná analýza principu a možností autostereoskopického vícehledového zobrazování:** zkoumání vlastností takového displeje a jejich matematický popis, odvození rozložení zón, ve kterých je možné dosažení neznehodnoceného prostorového vjemu, (kap. 4).
- Vyřešit problém vhodné **konfigurace kamer** tvořících stereo pár. Prioritou je vyhnout se složité mechanické kalibrovatelné soustavě a maximálně zjednodušit akviziční hardware, (kap. 5).
- **Návrh a vývoj nové metody, která vygeneruje další potřebné pohledy na libovolnou scénu** na základě definované prostorové konfigurace kamer a to bez explicitního vytvoření prostorového modelu, (kap. 5).
- **Vytvoření složeného obrazu** pro více než dvoupohledovou autostereoskopickou jednotku (princiálně popsanou v úvodu) s využitím reálně snímaných i generovaných pohledů jako vstupních dat.
- **Implementace a ověření stěžejních postupů** formou funkční aplikace umožňující ověření vlivu důležitých parametrů na výsledky autostereoskopického zobrazení.
- **Zefektivnění implementovaných algoritmů** na základě výsledků experimentů.

4 MINIMÁLNÍ VZDÁLENOST POZOROVATELE OD DISPLEJE

Tato část navazuje na kapitolu 1 popisující mimo jiné princip osmipohledového autostereoskopického displeje a zde je uveden vztah pro výpočet minimální vzdálenosti, ve které má pozorovatel ještě korektní prostorový vjem. Jestliže se totiž pozorovatel dostane do blízkosti obrazovky na takovou vzdálenost, kdy už paralaxa v obou krajních bodech obrazovky překročí do další osmice, vznikne v těchto oblastech linie přes obraz, která jej znehodnotí. Protože k pohybu pozorovatele ve vertikálním směru dochází pouze v omezené míře je uvažován pouze jeho horizontální pohyb. Pohyb pozorovatele před obrazovkou simuluje proměnná x , resp. její změna Δx , přičemž výchozí pozice s nulovou hodnotou je v ose obrazovky.

Uvažujeme-li horizontální rozměr zobrazovací plochy x , vzdálenost filtračního pole WLSFA od LCD panelu d a stranu jednoho kompletního čtvercového pixelu p , můžeme odvodit pro pozorování z jednoho bodu následující vztah

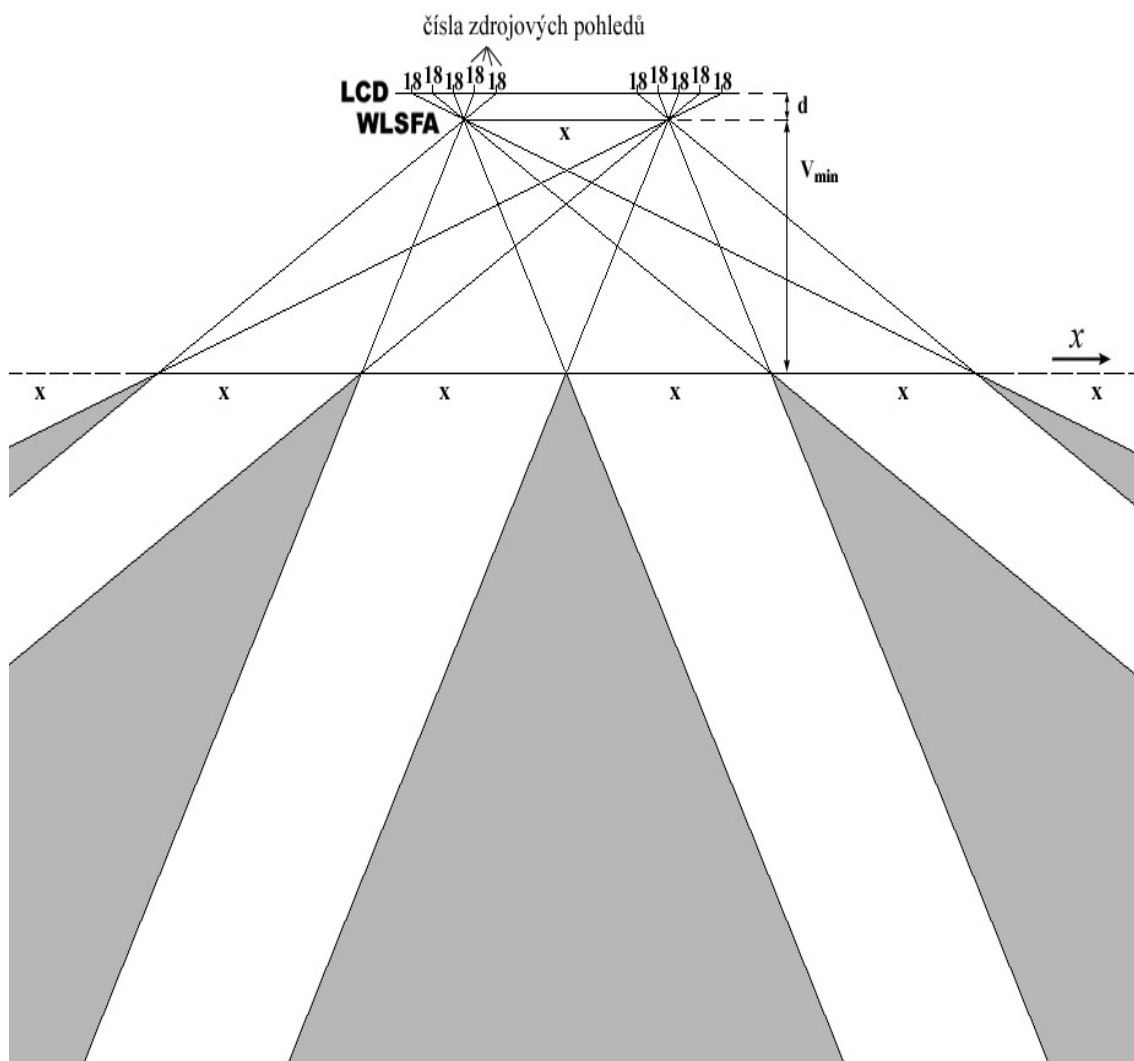
$$v_{\min} = d \left(\frac{x}{7s} - 1 \right), \quad (1)$$

kde $s = p/3$ je šířka subpixelu (jednoho barevného elementu) a v_{\min} značí nejmenší možnou kolmou vzdálenost jednoho pozorovacího bodu od roviny obrazovky. Při pohybu pozorovatele po přímce rovnoběžné s rovinou obrazovky a vzdálené od ní v_{\min} bude moci pozorovatel vidět neporušený obraz pouze v bodech kdy

$$\Delta x = k \cdot x, \quad \text{kde } k = 1, 2, 3, \dots \quad (2)$$

Hlavní nevýhodou tohoto principu zobrazování je velmi malé rozlišení. WLSFA filtr je v podstatě jemná síť, která kromě žádaného 3D efektu zapříčiňuje také drastické snížení rozlišení. Tento nepříznivý vliv je možné odstranit technologií, která zmenší rozměry subpixelů. Při současné úrovni miniaturizace se dá předpokládat, že tento úkol bude technologicky brzy vyřešen. Ze vztahu (1) mimo jiné vyplývá, že zmenší-li se rozměr subpixelu – konstanta s – zvětší se nutně minimální vzdálenost dosažení korektního prostorového vjemu v_{\min} . Tento nežádoucí účinek je však možné kompenzovat snížením vzdálenosti mezi WLSFA a LCD panelem označovaným jako d .

Na obr. 2 je principiální uspořádání autostereoskopického displeje s tmavě vyznačenými zónami korektních stereo párů. Aby mohl pozorovatel vnímat celý prostorový obraz bez rušení (vznikajícího překřížením zdrojových pohledů v oblasti svislého pruhu), musí se obě jeho oči nacházet v korektní zóně.



Obr. 2: Zóny korektních stereo párů.

V této kapitole byly představeny některé úvahy, které detailně analyzují princip a možnosti použitého autostereoskopického zobrazování. Tyto myšlenky byly publikovány se zahraničním ohlasem. Princip tohoto zobrazování je založen na relativně jednoduchých fyzikálních zákonech, avšak zpracování složeného obrazu, který je nutno vytvořit pro promítnutí na LCD panel za filtrem, vyžaduje složitější aparát spadající do oblasti číslicového zpracování obrazových dat specializovaného na počítačové vidění. Z popsanych principiálních vztahů vyplývá, že je možné hlavní nedostatek – malé rozlišení tohoto displeje – odstranit další miniaturizací rozměru barevných elementů. Vliv této úpravy na rozměry zón korektních stereo párů je kompenzovatelný změnou vzdálenosti filtru WLSFA od LCD panelu.

5 EFEKTIVNÍ METODA GENEROVÁNÍ NOVÉHO POHLEDU

5.1 JEDNODUCHÁ STEREO GEOMETRIE

Značného zjednodušení při hledání korespondencí ve stereo páru lze dosáhnout pomocí speciální konfigurace kamer. V případě, že budou osy obou kamer stereo dvojice paralelní už na úrovni hardware, mohou být na základě zavedena tato zjednodušení:

- Odpovídající si body musí mít identické souřadnice y .
- Pohyb (mapování) bodů vniklých promítnutím 3D scény do obrazových rovin referenčních snímků lze uvažovat pouze v jednom směru. Je-li zvolen levý obraz jako referenční, pak je při hledání korespondencí v pravém obraze možné uvažovat pouze pohyb pixelů levého obrazu zprava doleva.

Nechť $\mathbf{L}(x_L, y_L)$ a $\mathbf{R}(x_R, y_R)$ označují matice úrovní jasu jednotlivých pixelů levého, resp. pravého obrazu stereo páru, přičemž levý obraz je uvažován jako referenční. Pro většinu obrazových bodů (vyjma okludované) v levém obraze je možné nalézt protějšek v obraze pravém. V závislosti na jejich hloubce v reálné scéně (souřadnici Z) se tyto dvojice korespondujících bodů liší opět pouze v horizontální souřadnici, ovšem nyní vyjádřené v diskrétních hodnotách pořadí pixelu v řádku. Protože je uvažováno číslování souřadnic pixelů zleva doprava lze napsat

$$x_R = x_L - \mathbf{d}(x_L, y), \quad (3)$$

přičemž $y_L = y_R = y$ a $\mathbf{d}(x_L, y)$ značí dvourozměrné pole diskrétních hodnot disparit d jako celých kladných čísel (ve stejném formátu jako souřadnice pixelů). S rostoucí hloubkou bodu v reálném světě klesá disparita mezi promítanými body ve stereo páru, proto pole $\mathbf{d}(x_L, y)$ v podstatě odpovídá DEPTH matici a hodnoty d udávají poměrově vyjádřenou vzdálenost objektů ve scéně od kamery.

Je-li známa disparita, mohou být souřadnice mezilehlého, generovaného, pohledu umístěného mezi oběma zdrojovými vypočítány jako:

$$\begin{aligned} x_G &= x_L - d \cdot t \\ y_G &= y \end{aligned}, \quad (4)$$

kde $t \in \langle 0, 1 \rangle$ je transformační koeficient pro definici pozice nového pohledu. Nulová hodnota reprezentuje obraz sejmutý levou kamerou, číslo 1 reprezentuje pravý obraz a mezilehlé hodnoty v tomto intervalu spojitě udávají transformační koeficienty všech možných mezilehlých pohledů snímaných virtuální kamerou.

Vztah (4) je transformační vztah lineárního warpingu (deformace) levého snímku na základě disparitní informace získané ze stereo páru kamer konfigurovaných do jednoduché stereo geometrie. Dále budou uvažována pouze zdrojová obrazová data snímaná s dodržением podmínek nutných pro zpracování použitím tohoto vztahu.

5.2 URČENÍ DISPARITY

5.2.1 DSI (Disparity Space Image – disparitní prostor)

Pro nalezení disparit jednotlivých pixelů referenčního (levého) snímku je nutná kompletní analýza jasů všech pixelů ve stereo páru. Základem pro další zpracování bude struktura přehledně shrnující číselně vyjádřenou míru podobnosti, tedy v podstatě pravděpodobnost všech možných korespondencí. Při programové realizaci je vhodná počáteční alokace paměťového prostoru pro uložení těchto údajů: alokace 3D pole $\mathbf{DSI}(x_L, y, d)$. Je zřejmé, že toto pole bude mít svoje tři rozměry dány počtem řádků a sloupců zdrojových obrazů (ty musejí být identické) a velikostí maximální dovolené disparity d . Souřadnice x_L je použita protože levý snímek byl pro nalezení disparity zvolen jako referenční.

Existuje více možností výpočtu hodnot \mathbf{DSI} . Nejjednodušší, avšak nikoli nejbezzvýznamnější z těchto algoritmů jsou absolutní rozdíly:

$$\mathbf{DSI}(x_L, y, d) = \text{abs}(\mathbf{L}(x_L, y) - \mathbf{R}(x_L - d, y)). \quad (5)$$

Experimentální výsledky ukazují, že i tento relativně jednoduchý algoritmus má množství výhod, které jej upřednostňují před komplexnějšími algoritmy, které jsou mnohonásobně více výpočetně náročné. Finální generované pohledy založené na vztahu (5) ukazují, že úspora systémového času je při použití tohoto jednoduššího výpočtu spojena i s dobrými výsledky při generování nového pohledu. Dále je možné zahrnout do vyhodnocování podobnosti i určité okolí (okno) kolem souřadnice zpracovávaného pixelu, což může do určité míry přinést zpřesnění výsledků vyloučením nahodilých odchylek. Vedle výpočetní náročnosti takového algoritmu je nevýhodou ztráta detailů výsledné disparitní mapy při větších rozměrech okna.

Jednotlivé hodnoty v \mathbf{DSI} představují kvantitativně vyjádřenou míru s jakou aktuální pixel v \mathbf{DSI} se souřadnicí (x_L, y, d) v referenčním levém obraze koresponduje s pixelem se souřadnicemi $(x_L - d, y) = (x_R, y)$ v obraze pravém, který je uvažován jako vztažný. Toto může být aplikováno na každý pixel s výjimkou některých hraničních oblastí, kde je odpovídající obrazová informace v pravém obraze okludována.

Pro další zpracování takto naplněného \mathbf{DSI} je vhodné pracovat s jeho řezem, tedy s dvourozměrným polem $\mathbf{S}(d, x)$, které popisuje situaci korespondencí pro jednu dvojici řádků ve stereo páru obrazů. Ukázalo se, že pouhé nalezení minima v tomto řezu není ani pro určení přibližné disparitní mapy dostatečné. Je tedy nutné aplikovat vyspělejší metodu pro nalezení cesty přes takto připravená 2D pole jednotlivých řezů 3D pole \mathbf{DSI} . Tato metoda musí brát v úvahu nejenom hodnoty elementů, ale musí chápat tento problém komplexně jako nalezení „cesty nejmenšího odporu“.

Stav, kdy je \mathbf{DSI} naplněn, ale finální pole disparit ještě nebylo stanoveno, je vhodný pro úpravu hodnot \mathbf{DSI} za účelem zlepšení jeho vlastností pro aplikaci metody určující disparitní mapu. Pro metodu dynamického programování (jejíž

popis bude následovat), která je nezávisle aplikována na jednotlivé řádky ve stereo páru, je velmi důležitá konzistence sousedících hodnot. Může být realizována jako sumace hodnot rozdílů (případně s průměrováním) přilehlých hodnot v určitém okolí daného pixelu. Průměrování hodnot je sice logickým řešením problémů, kterými trpí prosté sečítání hodnot u okrajů zpracovávané matice, ale pro její zamýšlené využití není nutné tuto operaci provádět. Pro naprostou většinu elementů (mimo hranice) **DSI** by byla operace dělení pouze zbytečným plýtváním strojového času, jak ukázaly i experimenty. Ošetření tohoto vlivu by našlo opodstatnění až v případě, že by bylo pro zachování konzistence v **DSI** použito značně velké okolí. Pro okna v řádu jednotek pixelů však postačí pouze automatická změna parametru aditivní konstanty v základním algoritmu dynamického programování.

Výsledkem této funkce je vyhlazené 3D pole **DSI**, připravené pro další fázi. Obecně proti sobě stojí dva požadavky na stanovení disparity: požadavek **na spojitost** mapy (pro plynule se měnící hloubku scény) a požadavek na kvalitní detekci hran, čili **na nespojitost**. Z těchto protimluvů plyne, že exaktní určení hloubkového reliéfu není prakticky možné, existuje pouze hledání rovnováhy mezi oběma požadavky. Požadavek na spojitost mapy splňuje zmíněná aplikace vyhlazovacího filtru, zatímco detekci nespojitostí implementuje dynamické programování.

5.2.2 Dynamické programování

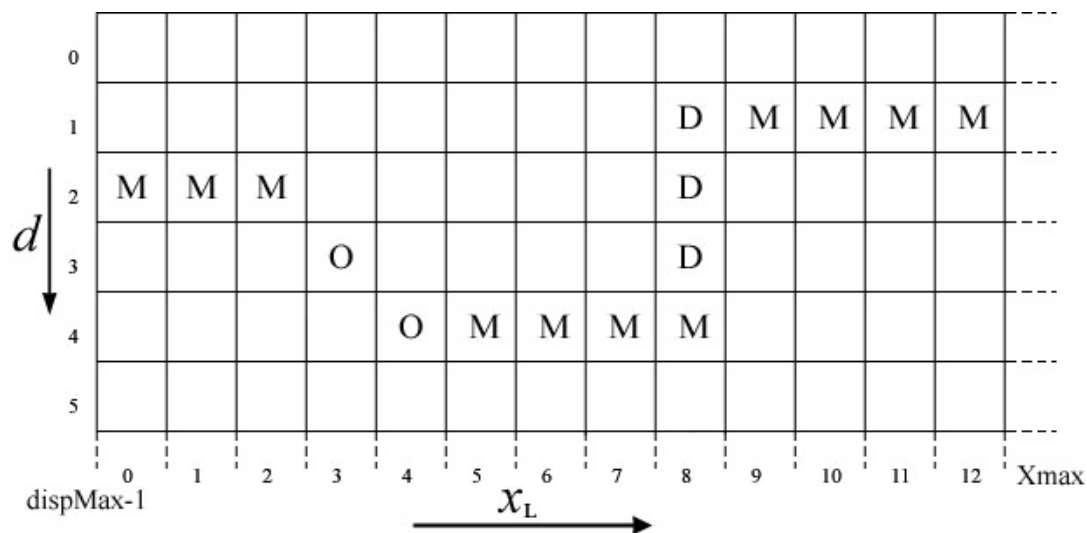
Poté, co byl **DSI** naplněn, je hledána efektivní metoda pro nalezení „cesty nejmenšího odporu“ přes hodnoty vyjadřující odlišnost všech korespondenčních alternativ pro dvojici řádků-přes 2D řez **DSI**. Za tímto účelem byla užitá metoda dynamického programování, což je obecně metoda redukce doby běhu programu nutné pro výpočet algoritmu pomocí tří kroků:

- rozložení daného problému na dílčí subproblémy,
- optimální řešení těchto subproblémů s použitím rekurze,
- finální použití těchto optimálních dílčích řešení ke konstrukci optimálního řešení originálního problému.

Cílem je rozdělení hlavního problému na co nejmenší množství jednoduše řešitelných subproblémů. Řešení korespondenčního problému stereo páru pomocí dynamického programování vychází z Needleman-Wunschova algoritmu přiřazení globálních sekvencí, který byl prezentován v roce 1970 jako metoda pro hledání podobnosti dvou proteinů pomocí hodnocení sekvencí jejich aminiokyselin [13]. V současné době se tento princip v různých modifikacích používá pro různé účely. Jednoduchou modifikací je možné využít tuto techniku i pro nalezení nejkratší cesty přes hodnoty připraveného pole 2D řezu **DSI**, pole $S(d, x)$. Výsledkem této metody je pro každý řádek disparitní reliéf s vyznačením částečně okludovaných pixelů. Příklad takového reliéfu je na obr. 3, kde jsou typy jednotlivých pixelů označeny počátečními písmeny jejich stavů:

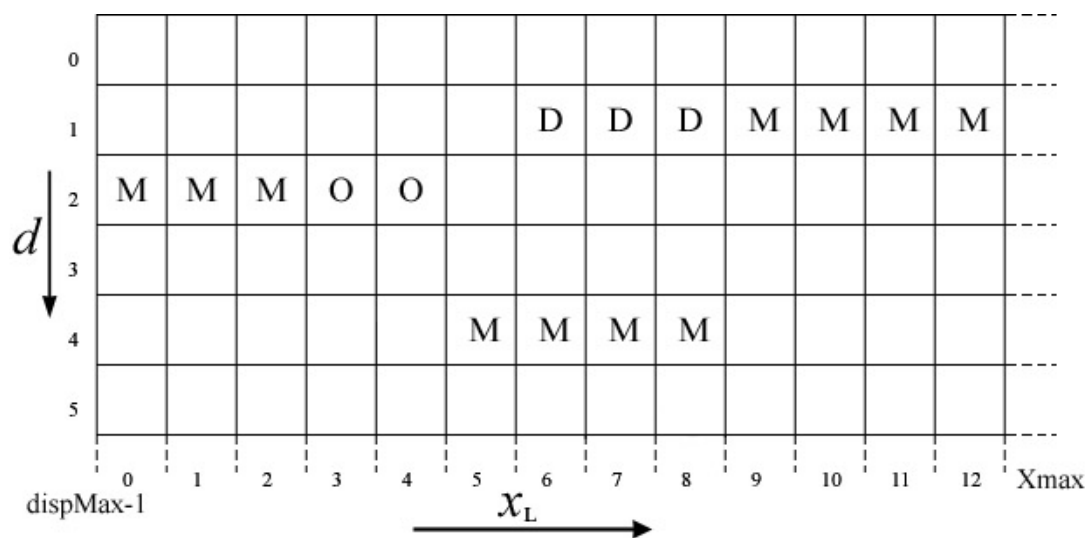
- *matched*, tj. správná korespondence,
- *occluded*, tj. překrytý, okludovaný pixel v referenčním (levém) obraze,

- disoccluded, tj. pixel odkrytý v referenčním (levém) obrazu.



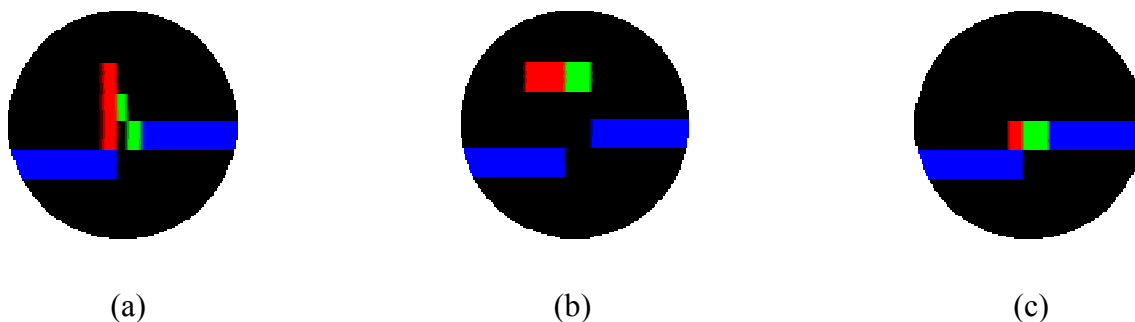
Obr. 3: Příklad disparitního reliéfu před úpravou nespojitosti.

Vzhledem k nespojitostem správně určených elementů v takovémto poli je však nutné ještě další zpracování, které zajistí správnou pozici okludovaných elementů. Příklad vlivu tohoto zpracování je na následujícím obrázku.



Obr. 4: Příklad disparitního reliéfu po úpravě nespojitosti.

Na obr. 4 je tedy výsledný disparitní reliéf. Při experimentech s různými testovacími scénami během vytváření této práce se však ukázalo, že může nastat stav ilustrovaný na obrázku 5 (a). Na tomto obrázku jsou jednotlivé stavy pixelů odlišeny barevně. Zeleně jsou označeny buňky ve stavu occluded, modře ve stavu matched a červeně ve stavu disoccluded.



Obr. 5: Speciální případ disparitního reliéfu: (a) před úpravou nespojitosti, (b) po úpravě nespojitosti bez ošetření tohoto případu, (c) po úpravě nespojitosti s ošetřením tohoto případu.

Takovýto reliéf může znamenat, že ve skutečné scéně existuje oblast, která není vidět ani z jedné kamery referenčního stereo páru, což se v disparitní mapě projeví jako „štěrbina“ v korektně označených buňkách, jak je vidět na obr. 5 (b). Scéna tohoto typu se však v oblasti zamýšlených disparit prakticky nevyskytuje, proto je takovýto reliéf chybný a je vhodné jej odstranit doplněním základního kódu. Po jeho ošetření pak bude podoba reliéfu vždy analogická k příkladu na obr. 5 (c).

Následující obrázky 6 až 8 demonstrují vliv ošetření tohoto reliéfu na kvalitu disparitní mapy. Na obr. 6 je nejdříve uveden levý referenční snímek jednoduché testovací scény se židlí, kde mohou být objekty vnímány intuitivně. Na obr. 7 jsou detaily, ve kterých se projevuje chybné určení reliéfu jako tmavé místo na disparitní mapě, i když je vidět, že takováto prohlubenina do daných oblastí jistě nepatří. Na obr. 8 je vidět v odpovídajících detailech odstranění nežádoucího vlivu speciálního reliéfu. (Vliv nežádoucího speciálního reliéfu na generované pohledy demonstrují obr. 12 a 13.)



Obr. 6: Levý referenční snímek testovací scény.



Obr. 7: Disparitní mapa získaná bez ošetření nežádoucího speciálního reliéfu.



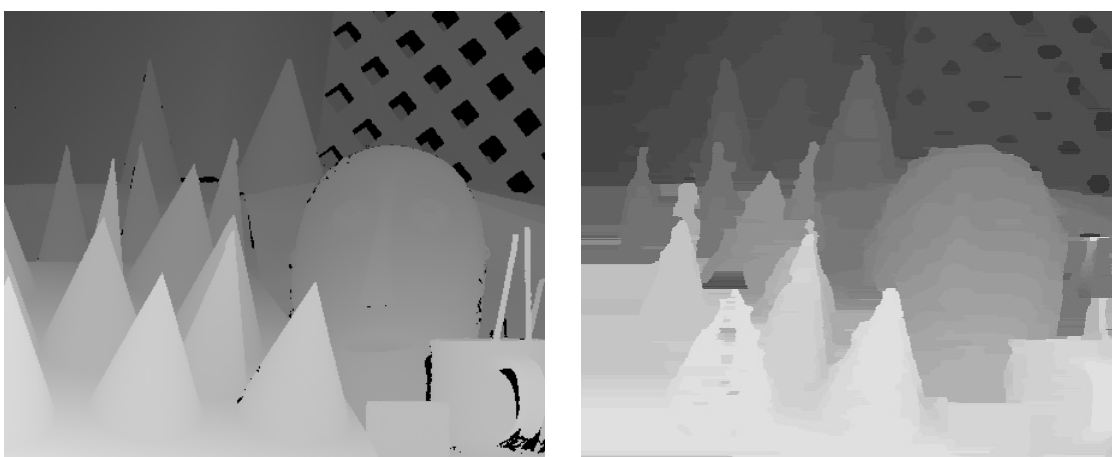
Obr. 8: Disparitní mapa získaná s ošetřením nežádoucího reliéfu.

5.2.3 Experimentální výsledky

V návaznosti na publ. [14] je možné porovnat výsledky vytvořeného programu s hloubkovou mapou získanou pomocí aktivního senzoru. Ve zmíněné literatuře je prezentován postup sejmutí skutečné hloubkové mapy, výsledky a zdrojová data jsou dostupná pro porovnání. Na obrázku 9 je tedy nejdříve testovací dvojice a následuje porovnání hloubkových map na obr. 10.



Obr. 9: Testovací dvojice pro aktivní senzor



(a)

(b)

Obr. 10: Porovnání disparitních map: (a) získaná aktivním senzorem, (b) vypočtená a normovaná.

Z tohoto porovnání vyplývá, že metoda dynamického programování s daným nastavením parametrů dosahuje poměrně dobrých výsledků už na úrovni klasické disparitní mapy (resp. DEPTH matice, protože rozdíl mezi disparitní mapou a DEPTH maticí je spíše filozofický). Jisté nedostatky je možné vidět v oblastech ostrých hrotů a tenkých předmětů, které jsou rozmazány či vymazány zejména díky použité filtraci.

5.3 GENEROVÁNÍ DALŠÍCH POHLEDŮ

V kapitole 5.1 byl popsán vztah pro výpočet souřadnic syntetizovaného pohledu (tzv. lineární warping). Při jeho implementaci je nutné vzít v úvahu, mimo jiné, možnost jasových odlišností ve zdrojové stereo dvojici. Na rozdíl od okludovaných buněk je nutné určit jasovou informaci obrazových elementů nového pohledu danou transformací buněk typu matched pomocí jasu (barvy) obou zdrojových snímků stereo páru. I v případě 100% korespondující dvojice pixelů je totiž možné, že se jejich jas liší díky různým světelným podmínkám při expozici obou obrazů. Plynulý přechod od jasu levého obrazu k pravému byl vyřešen rovnicí, která váhuje vliv příslušného pohledu na základě horizontální souřadnice v referenčním zdrojovém pohledu:

$$N(x_L - d \cdot t, y) = \frac{(x_{MAX} - x_L) \cdot L(x_L, y) + x_L \cdot R(x_L - d, y)}{x_{MAX}}, \quad (6)$$

kde N je matice úrovní jasu pixelů nového obrazu, x_{MAX} je maximální hodnota horizontální souřadnice a ostatní proměnné vycházejí ze vztahů (3) a (4).

Finální generovaný pohled je na obr. 11. Jedná se o nejhorší případ pohledu přesně mezilehlou virtuální kamerou. V oblastech ostrých hran a tenkých předmětů je patrný vliv chybného určení disparity diskutovaného u obr. 10.



Obr. 11: Syntetizovaný pohled z referenčního stereo páru z obr. 7.27.
(přesně mezilehlý => nejhorší případ)

Následující obrázky 12 a 13 demonstrují vliv ošetření nežádoucího speciálního reliéfu na kvalitu finálního generovaného pohledu z testovacího snímku z obr. 6. Na obr. 12 je patrné zkreslení jím způsobené (zejména nesouvislost svislých hran) a na obr. 13 je vidět příznivý vliv odstranění tohoto reliéfu.



Obr. 12: Mezilehlý pohled generovaný pomocí disparitní mapy bez ošetření nežádoucího speciálního reliéfu.



Obr. 13: Mezilehlý pohled generovaný pomocí disparitní mapy s ošetřením nežádoucího reliéfu.

Nyní už zbývá provést pouze závěrečný krok: poskládání potřebného počtu generovaných pohledů do složeného obrazu podle algoritmu daného konkrétní zobrazovací jednotkou a jeho prostorové zobrazení.

6 ZÁVĚR

Ve smyslu cílů stanovených v kapitole 3 byl proveden výzkum ve dvou základních etapách. První z nich, teoretická, spočívala v detailní analýze principu osmipohledové autostereoskopické jednotky. Druhou, rozsáhlejší částí byl vývoj nové metody akvizice 3D scény pro tento displej.

Z teoretické analýzy principu užití autostereoskopické zobrazovací jednotky vyplývají zejména možnosti zlepšení vlastností tohoto zobrazování spočívající ve zmenšení rozměru barevných elementů LCD panelu (případně užití jiného zobrazovacího zařízení). Tímto způsobem je možné snížit hlavní nevýhodu této autostereoskopické zobrazovací jednotky, a to sice malé rozlišení. Dále byla odvozena geometrie zón korektních stereo párů, ve kterých je možné dosažení nezhodnoceného prostorového vjemu. Výsledky tohoto teoretického výzkumu byly publikovány s ohlasem u zahraničních odborníků.

V hlavní a nejobsáhlejší části práce byl představen nový přístup k syntéze reálné scény snímané pouze dvěma běžnými kamerami pro její vyobrazení na osmipohledovém autostereoskopickém displeji pomocí dynamického programování. Tato metoda analýzy stereo páru obrazů vychází z jednoduché konfigurace dvojice kamer. Snímky jsou pořízeny dvojicí paralelních kamer, korespondenční problém je tak zjednodušen na hledání odpovídajících si pixelů vždy ve dvojici řádků referenční dvojice snímků. K řešení korespondenčního problému jednotlivých řádků je použito modifikace Needleman-Wunschova algoritmu přiřazení globálních sekvencí upraveného pro účely analýzy stereo páru obrazů. Chybovost tohoto přístupu, pramenící zejména z nezávislosti zpracování dvojic řádků v referenčních obrazech, je snížena sumací disparitního prostoru **DSI** ihned po jeho naplnění hodnotami popisujícími pravděpodobnost korespondence na základě absolutního nebo mocněného rozdílu příslušných pixelů. Výpočetní náročnost předzpracování **DSI** je výrazně snížena pomocí automatického přepočtu aditivní konstanty v dynamickém programování.

Při experimentech s vyvinutou metodou byla zjištěna příčina výskytu některých chyb v generovaných pohledech. Byla vyvinuta procedura pro detekci a odstranění jejich příčiny: nežádoucího speciálního reliéfu disparitní mapy, což zlepšilo kvalitu vypočítané mapy a tím i kvalitu syntetizovaných pohledů.

Výsledný řetězec zpracování, na jehož vstupu je stereo pár obrazů a na výstupu libovolný počet generovaných mezilehlých pohledů, odstraňuje nevýhodu vícepohledových autostereoskopických displejů oproti dvoupohledovým metodám stereoskopického zobrazování. Představená metoda může chybějící obrazy pro vícepohledovou jednotku vygenerovat s minimální aproximací na rozdíl od dosud publikovaných metod záznamu prostorové informace pro přenos v zamýšleném konceptu 3D televize formou klasické DEPTH matice a jednoho referenčního

snímku. Nová metoda především usnadňuje získávání hloubkového reliéfu, protože namísto nákladného aktivního senzoru měřícího vzdálenost snímaných předmětů používá klasické CCD kamery. Další výhodou je implicitní detekce a zohlednění částečných okluzí, které zmíněné metody řeší jednoduše aproximací. Zahnutí těchto faktorů do výsledného konceptu přináší změnu pohledu na přenos univerzální prostorové informace dvěma způsoby:

- Přenos stereo páru jako jediné informace o prostorovém uspořádání scény. V tomto případě může být na přijímací straně buď pouze dvoupohledová zobrazovací jednotka, nebo vícepohledová vybavená výkonným software pro generování potřebného počtu dalších pohledů.
- Přenos modifikované DEPTH matice jednoho referenčního snímku a doplňkové okludované informace. Modifikovaná DEPTH matice by musela obsahovat výsledky robustní analýzy stereo páru na vysílací straně včetně definice částečně okludovaných míst referenčních snímků. Doplňková informace o obsahu okluzí představuje velmi silně komprimovatelnou obrazovou informaci o malých oblastech neviditelných např. v pravém referenčním snímku. V této variantě zůstává na přijímací straně pro případ potřeby generování dalších potřebných pohledů pouze nepříliš náročný software.

Obě naznačené koncepce zachovávají kompatibilitu se standardní digitální televizí tak, že referenční snímek (například levý snímek ze stereo páru) plní současně funkci běžného 2D obrazu. V této oblasti je ještě velký prostor pro další výzkumy, protože bude nutné rozhodnout, který ze způsobů bude vhodný k praktickému užití.

Všechny popsané algoritmy a principy byly implementovány formou univerzální přehledné dialogové aplikace v programovacím jazyce C++. Během programování byl kladen důraz na nízkou výpočetní náročnost stěžejních částí kódu, avšak na některých místech jsou zcela jistě rezervy z důvodu univerzálnosti a názornosti cílové aplikace 3Dview, která měla sloužit zejména pro experimenty s metodou. Výsledky takových experimentů byly průběžně uváděny v textu práce. S udaným nastavením byla doba výkonu nejdůležitějších a současně nejnáročnějších funkcí dynamického programování přibližně 2,2 sekundy na PC Pentium 4, 2.81GHz (testovací barevný stereo pár měl rozlišení 512 x 384 pxl).

Podstatné části celé práce byly průběžně publikovány jednak v tuzemských časopisech a jednak na konferencích jak domácích, tak i na zahraničních.

LITERATURA

- [1] FEHN, C. Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV. *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems X*. Pages 93-104, San Jose, CA, USA, January 2004.
- [2] HARTLEY, R. I., ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004, ISBN 0521 54051 8.
- [3] FISHER, R. B. *CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision*. School of Informatics University of Edinburgh, dostupné na: <http://homepages.inf.ed.ac.uk/rbf/CVonline/>
- [4] X3D Technologies GmbH *X3D-Tools 2.2*. Jena, 2002
- [5] BIRCHFIELD, S. *An Introduction to Projective Geometry (for computer vision)*. Dostupné na: <http://robotics.stanford.edu/~birch/projective/>
- [6] WOODS, A., DOCHERTY, T., KOCH, R. Image Distortions in Stereoscopic Video Systems. *Proc. of SPIE Stereoscopic Displays and Applications '93*, pp. 36 - 48, (San Jose, CA, USA), Feb. 1993.
- [7] MCMILLAN, L. *An Image-Based Approach to Three-Dimensional Computer Graphics*. PhD thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1997.
- [8] D. SCHARSTEIN, View Synthesis Using Stereo Vision, *Lecture notes in computer vision; Vol. 1583*, Springer, 1999.
- [9] BALLARD, D. H., BROWN, C. M. *Computer Vision*. Prentice-Hall, Inc., New Jersey, 1982.
- [10] FAUGERAS, O., LUONG, Q. T., PAPADOPOULOU, T. *The Geometry of Multiple Images*. MIT Press, 2001, ISBN 0262062208.
- [11] KLETTE, R., SCHLUNS, K., KOSCHAN, A. *Computer Vision: Three-Dimensional Data from Images*. Springer-Verlag Singapore Pte. Ltd., 2001, ISBN 9813083719.
- [12] ZHANG, Z., DERICHE, R., LUONG Q.-T., FAUGERAS. O. A Robust Approach to Image Matching: Recovery of the Epipolar Geometry. *Proc. Int'l Symposium of Young Investigators on Information\Computer\Control*, pages 7-28, Beijing, China, February 1994.
- [13] NEEDLEMAN, S. B., WUNSCH, C. D. *A general method applicable to the search for similarities in the amino acid sequence of two proteins*, *Jrnl Molec. Biol.* 48 p443-453, 1970.
- [14] SCHARSTEIN, D., SZELISKI, R. High-Accuracy Stereo Depth Maps Using Structured Light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 195-202, Madison, WI, June 2003.
- [15] PRATA, S. *Mistrovství v C++*, Computer Press, a.s., BRNO, 2004, ISBN 80-251-0098-7.
- [16] ŘÍHA, K., HUJKA, P. Generation of Novel Views from Image Completed with Depth Information. *Research in Telecommunication Technology 2004*. RTT 2004. Prague, Czech republic, 2004, ISBN 80-01-03063-6
- [17] ŘÍHA, K. *3D zobrazovací jednotka na principu Wavelength-Selective Filter Array*. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), ISSN 1213-1539, 2004, roč. 2004, č. 15
- [18] ŘÍHA, K., HUJKA, P. *Epipolární geometrie*. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), ISSN 1213-1539, 2005, roč. 2005, č. 17
- [19] ŘÍHA, K. Virtual Camera View Creation Using Epipolar Geometry. In *Proceedings of the 5th International Conference of PhD Students, ENGINEERING SCIENCES II*. Miskolc (Hungary): University of Miskolc, 2005, p. 253 – 257. ISBN 963-661-679-5.
- [20] ŘÍHA, K. Epipolar Geometry for 3D Image Transformation. In *Proceedings of the Telecommunications and Signal Processing TSP - 2005*. BRNO: VUT Brno, 2005, s. 97 - 102, ISBN 80-214-2972-0.
- [21] ŘÍHA, K., VRBA, K. 3D Scene Acquisition for Autostereoscopic Display. In *Proceedings of the International Conference on Signals and Electronic Systems ICSES '06*, Lodz-Poland, 2006.
- [22] ŘÍHA, K., VRBA, K. Výpočet disparity stereo páru pomocí dynamického programování. In *Research in Telecommunication Technology 2006*. FEKT VUT v Brně, 2006.

CURRICULUM VITAE

Jméno: Ing. Kamil Říha
Narozen: 16.10.1978 v Novém Městě na Moravě
Kontakt: rihakamil@phd.feec.vutbr.cz

Vzdělání

- 2003 – 2006 Postgraduální studium oboru Teleinformatika v programu Elektrotechnika, elektronika, komunikační a řídicí technika na Fakultě elektrotechniky a komunikačních technologií Vysokého učení technického v Brně.
- 1998 – 2003 Magisterské studium v oboru Elektrotechnika a sdělovací technika na Fakultě elektrotechniky a informatiky Vysokého učení technického v Brně.
- 1993 – 1998 Integrovaná střední škola, Brno, Purkyňova 97, obor Provozní elektrotechnika.

Účast na řešení projektů

Rok 2004

- Statistické vyhodnocení rozměrů otvorů se subpixelovou přesností, IS1639.

Rok 2005

- Vícepohledové prostorové zobrazování z nekompletního zdroje obrazů, IS2839.
- Modernizace předmětu Multimediální služby, IS3098.

Rok 2006

- Morphing mezilehlých pohledů ze stereo páru, IS2428.
- Inovace laboratorních úloh navazujících na části přednášené látky v předmětu Elektroakustika, IS3142.

ABSTRAKT

Hlavním cílem této práce byl výzkum a ověření nové metody akvizice 3D scény pro její prostorové zobrazení. Na rozdíl od existujících principů autostereoskopického bariérového zobrazování, které používají více než dva pohledy snímání z různých (ekvidistantních) perspektiv, byla hledána metoda vytváření složeného obrazu pro takovou autostereoskopickou vícepohledovou jednotku bez nutnosti snímání a přenosu všech potřebných pohledů a bez vytváření explicitního prostorového modelu libovolné zobrazované scény, založená pouze na disparitní (prostorové) informaci obsažené ve stereo páru obrazů snímáných dvěma kamerami. Disparita je určena pomocí techniky dynamického programování.

ABSTRACT

The main goal of this work was the research and testing of novel 3D scene acquisition method for spatial displaying. In contrast to existing autostereoscopic barrier displaying approaches using more than two views captured from different (equidistant) perspectives, a method is sought for 3D autostereoscopic picture construction for such multi-view display without the necessity of capturing and transmission all the needed views and without creation an explicit spatial model of an arbitrary displayed scene, based only on disparity (spatial) information included in stereo pair. The disparity is computed using the dynamic programming technique.