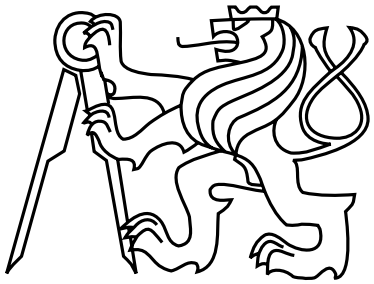
České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra počítačů



Diplomová práce

Pohledově závislá aplikace textur v reálném čase

Bc. Daniel Princ

Vedoucí práce: Ing. David Sedláček, Ph.D.

Studijní program: Otevřená informatika, Magisterský

Obor: Softwarové inženýrství

6. května 2014

Poděkování

Chtěl bych poděkovat vedoucímu této práce, panu Ing. Davidovi Sedláčkovi, Ph.D., za poskytnutí cenných rad a nápadů.

Prohlášení

Prohlašuji, že jsem práci vypracoval samostatně a použil jsem pouze podklady uvedené v přiloženém seznamu.
Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze, 12. 5. 2014 .....

Abstract

TBD

Obsah

Seznam obrázků

Seznam tabulek

Kapitola 1

Úvod

V dnešní době umožňuje počítačová grafika renderovat realistické obrazy 3D světa. Nicméně aby toto bylo možné, musí existovat 3D modely objektů, které chceme zobrazovat. Tradiční způsob, jak získat tyto modely, je jejich ruční vytváření v modelovacích programech. To je velmi pracný a zdlouhavý proces a jeho výsledek není dostatečně realistický. Proto se v současnosti vyvíjí postupy, jak přímo digitalizovat reálné objekty. Existují dva základní přístupy, které se o toto pokoušejí. Nejpřesnější metoda je pravděpodobně laserové skenování objektů. Druhou, mnohem levnější a dostupnější variantou, je rekonstrukce objektů z fotografií. Těmito problémy se zabývá počítačové vidění a částečně také počítačová grafika, do které spadá zobrazování rekonstruovaných modelů. Cílem 3D počítačové rekonstrukce je tedy co nejvěrněji převést reálné objekty do digitální podoby. Aplikaci najdeme ve virtuální realitě, online prohlídkách, počítačových animacích nebo v herním či filmovém průmyslu.

Tato práce se zabývá jednou částí 3D rekonstrukce a to texturování objektů z fotografií. Na vstupu očekáváme již zrekonstruovaný 3D model (síť trojúhelníků) a větší množství (desítky až stovky) kalibrovaných fotografií z různých úhlů. Textura společně s materiálem slouží k popisu povrchu a je důležitá pro vnímání barvy a detailní struktury povrchu [ŽBSF04]. Aplikace textury vede z výraznému zvýšení vizuální kvality objektu za relativně malou cenu. Často je vizuálně i výkonově jednodušší použít detailní texturu a jednoduchý model s menším počtem vrcholů.

Syntéza textur z fotografií je poměrně složitá, protože na výsledný model se díváme i z jiných pohledů, než ze kterých byly pořízeny fotografie. Problémy, které při tomto procesu vznikají jsou popsány zejména v následující kapitole. Cílem této práce je vytvořit aplikaci, která v reálném čase mapuje fotografie na model. Protože aplikace funguje v reálném čase, může zohlednit aktuální pohled virtuální kamery a na základě pozice kamery vybrat nejvhodnější fotografie pro vytvoření textury. K tomu využívá velkého výkonu současných GPU.

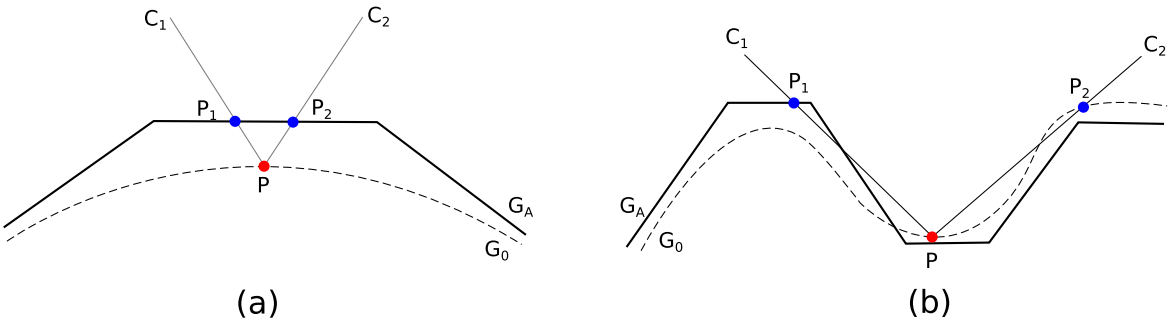
[[obrázek model s a bez textur TODO]]

Kapitola 2

Analýza problému a návrh řešení

2.1 Popis problému

Předpokládejme, že se díváme na statickou scénu s konstantním osvětlením, kde se pohybuje pouze pozorovatel. Za těchto podmínek můžeme popsat libovolnou fotografii pomocí pozice a orientace kamery. Pokud bychom vyfotili sférickou fotografii v každé možné pozici kamery, mohli bychom vyrenderovat kompletní scénu z libovolného pohledu. Kombinací všech 3D pozic kamery (x, y, z) a směrů (θ, ϕ) získáme plenoptickou funkci $\mathbf{L}(x, y, z, \theta, \phi)$ [AB91]. Snahou je aproximovat funkci $\mathbf{L}$ pomocí konečného množství diskrétních vzorků (x, y, z, θ, ϕ) a z této reprezentace efektivně renderovat nové pohledy (s pomocí 3D geometrie). Povrch geometrie můžeme popsat jako funkci $\mathbf{G} : (x, y, z, \theta, \phi) \rightarrow (x_0, y_0, z_0)$, tedy jako mapování pohledových paprsků na 3D souřadnice povrchu. Jako $\mathbf{G}_0$ si označíme funkci skutečného povrchu, $\mathbf{G}_A$ reprezentuje aproximovaný povrch (rekonstruovaný 3D model), viz obr. 2.1.



Obrázek 2.1: (a) chyby v geometrii způsobují chybnou projekci, bod P na původní geometrii je promítnut na body P_1 a P_2 na aproximované geometrii. (b) chyby ve viditelnosti, bod P je chybně viditelný z kamery C_2 a naopak není viditelný z kamery C_1.

Důležitým prvkem je projekční matice kamery, která nám určuje projekční mapování 3D bodu (x, y, z) do 2D souřadnic (s, t) v i-té fotografii $\mathbf{P}_i : (x, y, z) \rightarrow (s, t)$, viz sekce 2.3.1. Z promítnuté pozice v obrázku pak můžeme přiřadit 3D bodu barvu $\mathbf{I}_i : (s, t) \rightarrow (r, g, b)$. Poté libovolný nový pohled $\mathbf{I}^V$ z virtuální kamery V můžeme vyjádřit pomocí rovnice 2.1 [EDDM+08].

$$\mathbf{I}^V(x,y,z,\theta,\phi) = \sum_i \mathbf{I}_i^V(x,y,z,\theta,\phi)\ \omega_i \tag{2.1}$$

kde

$$\mathbf{I}_i^V(x,y,z,\theta,\phi) = \mathbf{I}_i(\mathbf{P}_i(\mathbf{G}_A(x,y,z,\theta,\phi))) \tag{2.2}$$

$$\omega_i = \delta_i(\mathbf{G}_A(x,y,z,\theta,\phi))w_i(x,y,z,\theta,\phi) \tag{2.3}$$

a $\sum_i \omega_i = 1$. Notace $\mathbf{I}_i^V$ značí vyrenderovaný obrázek z pohledu V promítnutím vstupní fotografie $\mathbf{I}_i$ jako textury na povrch $\mathbf{G}_A$. δ_i určuje viditelnost a je 1, pokud je bod na povrchu $\mathbf{G}_A$ viditelný v kameře i a 0 pokud není. w_i je funkce, která určuje váhu kamery i pro každý paprsek. Rovnice 2.1 se snaží reprezentovat plenoptickou funkci jako lineární kombinaci promítnutých fotografií.

Zásadní problém spočívá v tom, že $\mathbf{G}_A \neq \mathbf{G}_0$, tedy již vstupní data vztahu 2.2 jsou zatížena chybou (obr. 2.1). Problémy přináší také nepřesná kalibrace kamery $\mathbf{P}_i$, což způsobuje další nepřesnosti při mapování.

2.2 Související práce

Existuje několik základních metod, jak mapovat fotografie z více pohledů na model. Jednou z možností je projekce všech fotografií a jejich následné sloučení pomocí nějakého váženého průměru, jako např. v [BMR01]. Nevýhodou takové metody je zejména vznik artefaktů ve výsledné textuře, často se projevuje tzv. ghosting, kdy se v textuře objeví několik kopií jednoho obrazce. Další variantou je vytvoření atlasu textur [APK08], kdy každá část modelu dostane svojí texturu z unikátního pohledu. Tato varianta je využita např. v [AMK10]. Tento přístup má nevýhodu ve vzniku švů na okrajích jednotlivých částí textur, které je pak nutné odstraňovat [LI07]. Kromě artefaktů je u těchto přístupů častým problémem rozostření některých částí textur. Většina zmiňovaných problémů vzniká kvůli nepřesné kalibraci fotografií (špatnému odhadu radiální distorze či ohniskové vzdálenosti) nebo nepřesně rekonstruovaným modelům. Tyto nepřesnosti jsou obvyklé, i když jsou dnes již algoritmy pro 3D rekonstrukci velmi kvalitní. Získat velmi přesně kalibrované fotografie je časově náročný proces a někdy i téměř nerealizovatelný, např. ve venkovních scénách.

V článku [AMK10] jsou uvažovány nepřesně vytvořené modely, které je nutné otexturovat z původních fotografií, i když na model přesně nepasují. Navrhují podle modelu upravit původní fotografie a z nich následně vytvořit atlas textur. Úpravu provádí tak, že ve fotografiích identifikují významné prvky, které naleznou na vytvořeném modelu a zpětně je promítají do původních pohledů. Poté deformují všechny fotografie, aby co nejlépe odpovídaly nepřesným modelům. Touto metodou se nezbaví všech artefaktů, ale omezí jejich výskyt.

Podobný přístup je použitý v [TM09], kde je popsán způsob, jak dynamicky deformovat několik textur najednou podle 3D modelu a poté je pomocí vážených faktorů sloučit dohromady na základě pozice virtuální kamery.

Odlišný přístup je použitý v článku [CCCS08]. Zde řeší texturování hustých modelů s jednotkami až desítkami milionů trojúhelníků. Pro takto husté modely nepoužívají textury, ale barvu přiřazují pouze vrcholům, což vzhledem k počtu vrcholů v modelu poskytuje dostatečné detaily. Základní princip algoritmu je takový, že model je vykreslen z pohledu jednotlivých kamer a poté jsou viditelné vrcholy promítnuty zpět do původních fotografií. Pokud je jeden vrchol vidět na více fotografiích, je použita funkce, která vybere nejvhodnější barvu pro daný vrchol. Tato funkce používá vážené masky, které jsou vygenerovány pro každou fotografii a udávají kvalitu jednotlivých pixelů. Pro určení kvality pixelu je použito několik různých metrik, pro každou fotografii je tedy vytvořeno více masek (každá maska má rozlišení stejné jako fotografie). Použité metriky jsou následující:

- Úhlová metrika - nejjednodušší metrika, která porovnává směr ke kameře s normálou plochy. Největší váha je, pokud jsou oba směry stejné.
- Hloubková metrika - váha pixelu je větší, pokud je povrch blíže ke kameře.
- Hraniční metrika - tato maska udává, jak daleko je pixel od okrajů fotografie a silulety v hloubkové mapě. Čím dále je pixel od okrajů, tím je jeho kvalita lepší.

Výsledná váha pixelu je získána vynásobením hodnot v jednotlivých maskách. Tím je zaručeno zachování lokálních minim v každé masce, což pomáhá odstraňovat pixely, které jsou v libovolné masce považovány za velmi špatné. Výsledná barva pro každý vrchol se získá porovnáním masek u všech fotografií, ze kterých je daný vrchol viditelný, a následným vybráním nejlepšího pixelu.

Výhoda tohoto způsobu je, že se nejedná o výpočetně složité operace, většina výpočtů je prováděna nad fotografiemi a není závislá na složitosti modelu. Další výhodou je možnost určit kvalitu jednotlivých fotografií podle maximální či průměrné kvality masky. Tím je možné některé nevhodné fotografie automaticky eliminovat a zrychlit celý proces. Nevýhodou této metody je nutnost hustých modelů, u modelů s nižším počtem vrcholů by výsledky této metody nebyly příliš kvalitní. Další problém nastává, pokud je rozlišení fotografií vyšší než rozlišení modelu (jeden vrchol se mapuje na více pixelů ve fotografiích), to vyžaduje další zpracování dat a zvyšuje složitost problému.

Všechny tyto algoritmy mají společné to, že z původních pohledů předem vytvoří texturu či atlas textur spojením všech fotografií, přičemž se snaží minimalizovat vznik artefaktů nebo případně vzniklé artefakty odstraňovat. Metoda navržená v této práci se zásadně liší tím, že žádnou takovou texturu nevytváří, ale pro každý aktuální pohled virtuální kamery vybírá množinu fotografií z nejvhodnějších pohledů a z této množiny vybírá nejvhodnější texely. Zásadní nevýhodou tohoto přístupu je velký výpočetní výkon, který je potřeba při zobrazování modelu. Tuto nevýhodu se snažíme minimalizovat efektivním využitím GPU.

Tato myšlenka není úplně nová, na podobném principu je založena např. metoda plovoucích textur [EDDM+08]. Tento algoritmus používá adaptivní nelineární metodu, která opravuje lokální nezarovnání textur vůči 3D modelu. K tomu určuje optický tok [1] mezi promítanými fotografiemi a příslušné textury kombinuje.

[1] v orig. optical flow

Navrhovaný postup využívá kombinaci lineární interpolace a odhadu optického toku. Nejprve je provedena projekce fotografií I_i na model z původních pohledů a scéna je vyrenderována z aktuálního pohledu virtuální kamery V. Tím vzniknou dočasné textury I_i^V. Poté je na jednotlivé páry textur I_i^V aplikován odhad optického toku, čímž vzniknou pole $W_{I_i^V \to I_j^V}$. Pro více než dvě vstupní fotografie je nutné provést lineární kombinaci vytvořených polí a sloučit je do výsledné textury I^v_{Float}. To je poměrně náročná operace, pro n vstupních fotografií je nutné vytvořit $O(n^2)$ polí. S tím se vyrovnávají v článku tím, že používají pro každý pohled jen 3 nejbližší fotografie. Plovoucí textury porušují epipolární geometrii, což umožňuje texturám kompenzovat nekvalitní kalibraci kamery a nepřesně zrekonstruované 3D modely.

Dále je nutné vypořádat se s vlastním zastíněním částí modelu, což je velmi běžná situace. K tomu je využita jemná mapa viditelnosti, která pro každý pixel určuje, zda je z dané kamery viditelný nebo ne. Oproti tradičním postupům, které obvykle využívají pouze hodnoty 1 a 0 (je nebo není vidět), je zde použita metoda, která nastavuje hodnotu z intervalu $(0, 1)$ pixelům, které jsou blízko hranic zastínění. Tím jsou odstraněny ostré hrany podél zastíněných částí, které jsou velmi často nepřesné a snižují výslednou vizuální kvalitu.

Na principu výběru nejlepší fotografie na základě aktuálního pohledu je založená metoda v článku [DTM96]. Při mapování jedné fotografie navrhují použití image-space stínových map pro řešení viditelnosti, protože to umožňuje efektivní implementaci pomocí z-bufferu. Při mapování více fotografií na model vybírají pro každý pixel vždy takovou fotografii, která se na daný povrch dívá pod nejlepším úhlem. To samozřejmě přináší viditelné švy, protože sousední pixely mohou pocházet z různých fotografií. Toto řeší zjemněním přechodů pomocí váženého průměru, váhu je určena podle rozdílu úhlu aktuálního pohledu a pohledu původní kamery. Dále pro lepší výsledky mají pixely na okraji fotografie menší váhu, čímž se snaží ještě více eliminovat vznik švů.

Dále navrhují jednoduchý algoritmus pro odstranění nežádoucích objektů, které se mohou vyskytovat na zdrojových fotografiích - např. auto či chodci před budovou, kterou chceme otexturovat. Uživatel může ručně vymaskovat tyto objekty předem zvolenou barvou a tyto pixely dostanou při texturování nulovou váhu a budou použita data z jiné fotografie. Pokud nejsou dostupná žádná data, vyplňují vzniklé mezery pomocí syntézy obrazu.

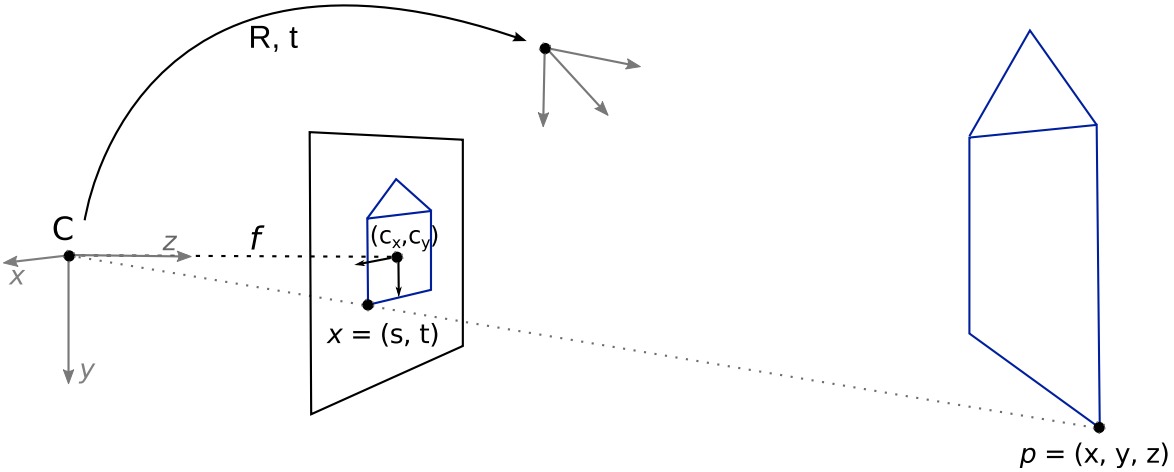
2.3 Návrh řešení

Algoritmus se skládá ze tří základních kroků - projekce fotografií na model, výběr nejlepších fotografií pro pokrytí celého modelu a výběr nejvhodnějšího pixelu z více fotografií pro otexturování fragmentu[2].

2.3.1 Projekce fotografie na model

Mapování jedné fotografie na model můžeme popsat jako projekci $P : (x, y, z) \to (s, t)$, tedy jako projekci 3D bodů modelu do 2D souřadnice ve fotografii, viz obr. 2.2. Nejběžněji se v počítačové grafice a vidění používá perspektivní projekce, která promítá 3D body p na

[2]Fragmentem je v textu vždy myšlena plocha na 3D modelu, který odpovídá jednomu výslednému pixelu na obrazovce. Stejně, jako je termín fragment používán v OpenGL.



Obrázek 2.2: Obrázek znázorňuje situaci při projekci 3D bodu na 2D souřadnice.

2D body x vydělením jejich z souřadnicí [Sze10]. V homogeních souřadnicích má kanonická perspektivní matice $\mathbf{P}_0$ jednoduchou formu:

$$x = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{P}_0} p \tag{2.4}$$

Po promítnutí 3D bodu projekční kamerou je nutné transformovat souřadnice na základě vlastností senzoru a orientace kamery vzhledem k počátku souřadnicového systému. *Kalibrační matice* $\mathbf{K}$ transformuje kanonickou perspektivní kameru na standardní projekční kameru $\mathbf{P}$:

$$\mathbf{P} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \mathbf{K}\mathbf{P}_0 \tag{2.5}$$

kde f je ohnisková vzdálenost v pixelech a bod (c_x, c_y) je optický střed vyjádřený v pixelech, obr. 2.2. Toto je zjednodušená matice kamery $\mathbf{K}$, která uvažuje senzor kolmý vůči optické ose a stejnou ohniskovou vzdálenost v osách x a y (což je v praxi nejběžnější varianta). Orientaci kamery vůči počátku souřadnicového systému definujeme pomocí 3×3 rotační kamery $\mathbf{R}$ a vektoru $\mathbf{t}$, čímž získáme výslednou 3×4 *matici kamery*:

$$\mathbf{P} = \mathbf{K} \, [\, \mathbf{R} \mid \mathbf{t} \,] \tag{2.6}$$

která provádí mapování bodu p_w v 3D světových souřadnicích do 2D souřadnic x ve fotografii:

$$x = \mathbf{P} \cdot p_w \tag{2.7}$$

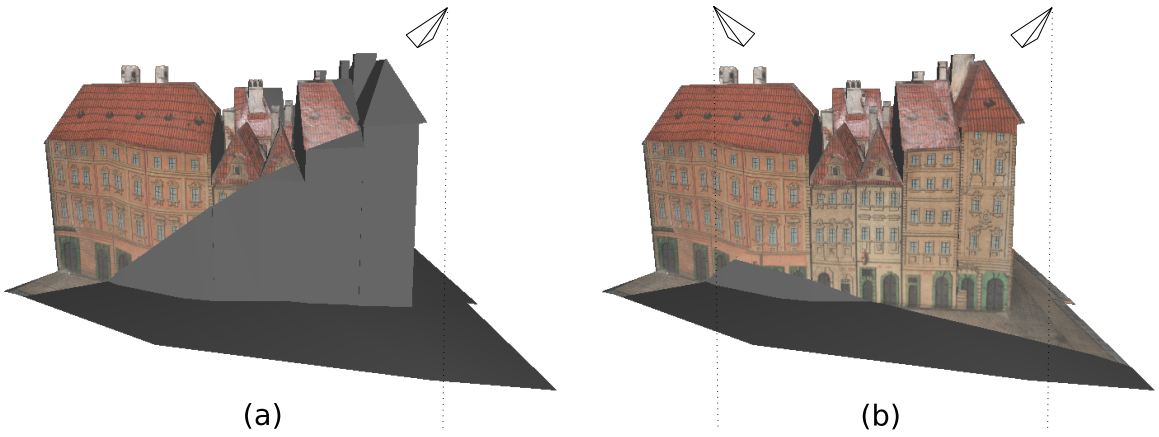
Existuje celá řada algoritmů pro nalezení matice kamery [HZ04], to ale není součástí této práce, kalibrované kamery jsou poskytnuté na vstupu společně s rekonstruovaným modelem a fotografiemi.

Při mapování fotografie na model je nutné brát v úvahu, že některé části modelu mohou být vzhledem ke kameře zastíněné a nemohou být z dané kamery správně otexturovány. Toto se nejčastěji řeší pomocí stínových map [SD02], které se předem vytvoří pro každou kameru. Aby byla tato metoda kvalitní, musí mít stínové mapy dostatečné rozlišení. Vzhledem k tomu, že aplikace bude využívat až stovky různých kamer, bylo by použití této metody paměťově velmi náročné.

Další možností je použít algoritmus vrhání paprsků, kdy se z kamery vrhají na scénu paprsky a podle zásahů s modelem se určí, které fragmenty jsou z kamery viditelné. Tato metoda také není příliš vhodná, protože aplikace bude v reálném čase využívat až desítky různých kamer zároveň a vrhání dostatečného množství paprsků z každé kamery by bylo výpočetně příliš náročné. Proto se jako nejlepší varianta se jeví využít jednodušší algoritmus, který spoléhá na seřazení kamer podle podobnosti se směrem virtuální kamery. Kvůli velkému množství vstupních kamer se dá očekávat, že pro většinu možných pohledů virtuální kamery bude existovat kamera s velmi podobným pohledem. Pro takovou kameru bude existovat jen velmi malé množství oblastí, které budou z kamery zastíněné, ale z virtuální kamery budou viditelné. Při texturování se kamery seřadí od "nejlepší" a každá kamera bude využita k otexturování oblasti, která ještě není pokrytá předchozí kamerou. Tím se výrazně omezí oblasti, které jsou otexturovány zastíněnou kamerou, ale zároveň jsou viditelné z virtuální kamery. K tomuto pravidlu je navíc zaveden práh, který určuje minimální úhel mezi směrem kamery a normálou plochy, pro který může být daná plocha kamerou otexturována.

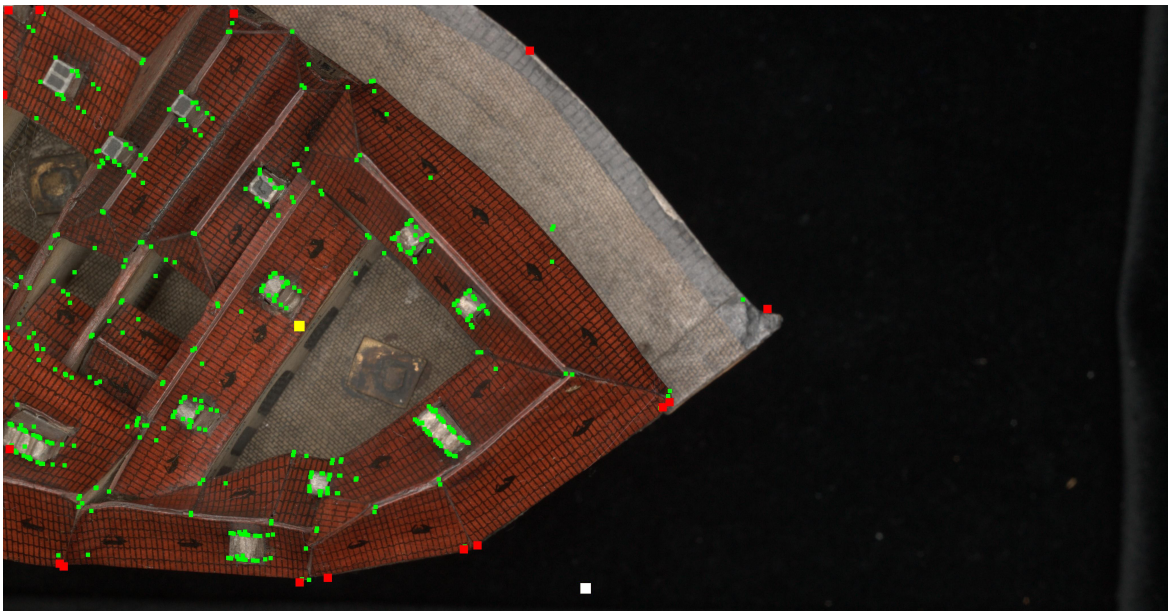
2.3.2 Výběr vhodných fotografií pro otexturování modelu

Téměř vždy nestačí k otexturování aktuálního pohledu pouze jedna fotografie. Proto je potřeba použít více fotografií pro vyrenderování modelu z nového pohledu. Na obrázku 2.3 je zobrazen model s jednou a se dvěma namapovanými fotografiemi. Při mapování více fotografií zároveň je nutné vyřešit dva základní problémy - které fotografie budou vybrány a jak bude fragment otexturován, pokud je jeho polohu možné promítnout do více kamer.



Obrázek 2.3: Projekce jedné (a) a dvou (b) fotografií na 3D model. Pozice kamer jsou pouze ilustrační.

Výběr fotografií je je prováděn na základě shody aktuálního pohledu virtuální kamery s pohledy vstupních kamer. Pro porovnání je nutné určit, kterým směrem jsou kamery natočeny. U vstupních kamer jde tento směr zjistit přímo z matice kamery, ale nemusí to být vždy výhodné. Pokud se kamera nedívá přímo na objekt (optická osa neprochází objektem), může být směr kamery poměrně zavádějící. Proto navrhujeme algoritmus, který tento směr koriguje na základě viditelné části modelu na fotografii. Tato metoda funguje tak, že se vrcholy rekonstruovaného modelu promítnou do fotografie, body ležící mimo fotografii se zahodí. Poté je nalezena konvexní obálka [And79] promítnutých bodů, která přibližně ohraničuje oblast modelu zobrazenou na fotografii, viz obr. 2.4. Z bodů ležících na konvexní obálce se spočítá centroid a nalezne se směrový vektor ze středu kamery procházející centroidem. Nalezený vektor se použije jako výsledný směr pohledu kamery. Algoritmus není příliš náročný, promítnutí n vrcholů do obrázku je provedeno se složitostí $\Theta(n)$, konvexní obálku ve 2D lze nalézt se složitostí $0(n\ log(n))$. Během tohoto algoritmu je zároveň vypočtena plocha konvexní obálky (plocha konvexního polygonu s vrcholy v bodech konvexní obálky). Tato plocha je později při texturování fragmentů zahrnuta do výpočtu váhy fotografie, viz sekce 2.3.3.



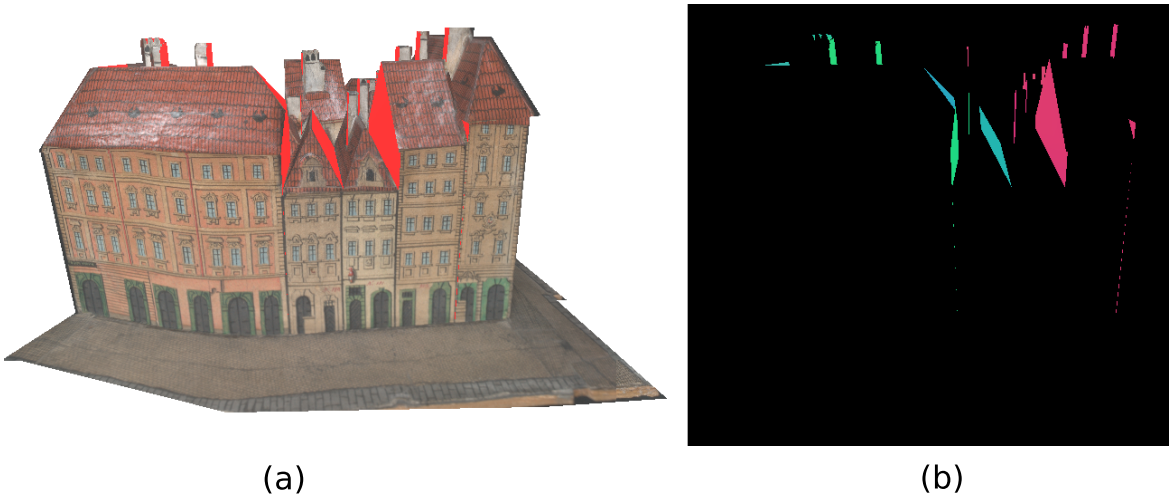
Obrázek 2.4: Výřez horní poloviny fotografie použité k texturování. Bílý bod označuje průnik optické osy s obrázkem, žlutý bod označuje vypočtený centroid. Ostatní body jsou vrcholy 3D modelu promítnuté do obrázku, červeně zvýrazněné vrcholy leží na konvexní obálce.

Podobná situace nastává při určování směru virtuální kamery. Pokud se díváme zvenku na model a otáčíme virtuální kamerou, vidíme model stále ze stejného směru a chceme, aby byl otexturován stejně. Proto není vhodné použít přímo směr, kterým se dívá virtuální kamera, ale spíše spojnici mezi pozicí kamery a modelem. Pro určení pozice modelu je použit centroid všech vrcholů. Odlišná situace nastává, pokud se virtuální kamera nachází uvnitř modelu. V tomto případě naopak musíme brát v úvahu přímo směr pohledu virtuální kamery, protože při otočení kamery vidíme jinou část modelu. K jednoduchému určení, zda je kamera

uvnitř nebo vně modelu, je možné použít osově zarovnaný ohraničující kvádr (AABB).

Po vypočtení směru pohledu vybereme nejlepší fotografie porovnáním se směry všech vstupních kamer. Počet fotografií potřebných k otexturování je závislý na složitosti scény a i na samotných fotografiích, proto se tato hodnota nenastavuje automaticky, ale uživatel jí musí zadat na vstupu.

Pro otexturování modelu často nestačí vybrat pouze fotografie, které nejlépe odpovídají aktuálnímu pohledu. Takové fotografie dobře pokryjí plochy, které jsou na aktuální pohled přibližně kolmé. Oproti tomu velmi špatně pokryjí plochy, které jsou téměř paralelní se směrem aktuálního pohledu, ale zároveň jsou ještě dobře viditelné, obr. 2.5 (a). K určení takových ploch je nutné zavést práh, který definuje maximální úhel mezi normálou plochy a směrem kamery, kdy je ještě možné plochu kvalitně otexturovat. Během testování se ukázalo vhodné použít práh přibližně 70°.



Obrázek 2.5: (a) Model otexturovaný šesti fotografiemi z kamer, které nejlépe odpovídají pohledu virtuální kamery. Červeně jsou zvýrazněny plochy, které nejde z kamer kvalitně otexturovat. (b) Vyrenderovaná textura s uloženými normálami v prvním průchodu algoritmu ze stejného pohleduj jaklo (a).

K efektivnímu nalezení těchto ploch navrhujeme dvou průchodový algoritmus. Nejprve se vybere daný počet fotografií, které nejlépe odpovídají aktuálnímu pohledu. Poté se v prvním průchodu model vykreslí a zjistí se, které plochy je s daným prahem možné z vybraných fotografií otexturovat. Pokud fragment není možné pokrýt žádnou fotografií, vykreslí se do textury aktuální normála. V prvním průchodu tedy vznikne textura, která obsahuje normály ploch, které je nutné pokrýt, obr. 2.5 (b). Poté je provedeno klastrování těchto normál a jsou určeny nejvýznamnější směry. Tento přístup je vhodný, protože klastrování jde efektivně provést na GPU bez nutnosti přenosu velkého množství dat mezi GPU a CPU. Poté se vyberou nové fotografie, které nejlépe odpovídají nalezeným směrům (kamery, jejichž směr pohledu je opačný k nalezeným normálám). V druhém průchodu se poté provede výsledný rendering.

2.3.3 Texturování z více fotografií

Během renderingu je nutné vyřešit, jak budou otexturovány fragmenty, pro které je možné použít data z více fotografií. Princip je velmi podobný, jako při výběru fotografií, ale snahou při texturování je také omezení vzniku švů a omezení chybného otexturování zastíněných oblastí. Během texturování se již používají pouze fotografie, které byly vybrány v předchozím kroku.

Pro každou fotografii je vypočtena váha, fotografie s nejvyšší váhou je použita pro otexturování. Jeden fragment je otexturován pouze z jedné fotografie, není použito žádné vážené míchání barev z více fotografií. To vede k tomu, že dva sousední pixely mohou být vybrány z různých fotografií a tím mohou vznikat viditelné švy. Vážené míchání barev, jako např. v [DTM96] vede k rozmazání textur a na testovacích datech k vizuálně horšímu výsledku.

Váha w_i i-té fotografie je složena z několika faktorů. V úvahu je brána shoda směru kamery d_c se směrem virtuální kamery d_v, stejně jako při výběru fotografií. Dále se uvažuje úhel mezi směrem kamery a normálou plochy N_f. Tím se preferují kamery s co nejvíce kolmým pohledem na danou plochu a zároveň je pravděpodobnější, že celá plocha bude otexturována z jedné fotografie. Nakonec je do výpočtu zahrnuta velikost plochy s_i, kterou na fotografii zabírá model, viz výpočet konvexní obálky v sekci 2.3.2. Celkový výpočet váhy shrnuje následující rovnice:

$$w_i = dot(d_v, d_c) \times \left(\frac{1 + dot(-N_f, d_c)}{1 - cos(\alpha)} + 1 \right) \times s_i \times \delta(i) \tag{2.8}$$

$$\delta(i) = \begin{cases} 1 & \text{pokud } \exists P_i : (x, y, z) \rightarrow (s, t) \\ 0 & \text{v opačném případě} \end{cases} \tag{2.9}$$

kde *dot* značí skalární součin dvou vektorů a α je limitní úhel, který určuje, zda je plochu možné z dané kamery otexturovat.

Musíme brát v úvahu, že máme dvě rozdílné množiny fotografií. Nejprve jsme vybrali fotografie, které nejlépe odpovídají aktuálnímu pohledu. Poté jsme nalezli doplňující fotografie, které pokryjí zbývající plochy. Proto musí rendering probíhat ve dvou částech. Nejprve se aplikuje původní množina fotografií. Až poté se použijí doplňující fotografie, které mohou otexturovat pouze fragmenty, které se nepodařilo otexturovat v první části. Tímto se jednak preferují fotografie blízké aktuálnímu pohledu a zároveň se omezí situace, kdy by doplňující fotografie mohly chybně otexturovat plochy, které jsou z jejich pohledu zastíněné. K tomu i tak může dojít, ale pouze ve velmi omezených částech výsledného renderu, vizuálně to tedy nebude příliš výrazné.

Kapitola 3

Implementace

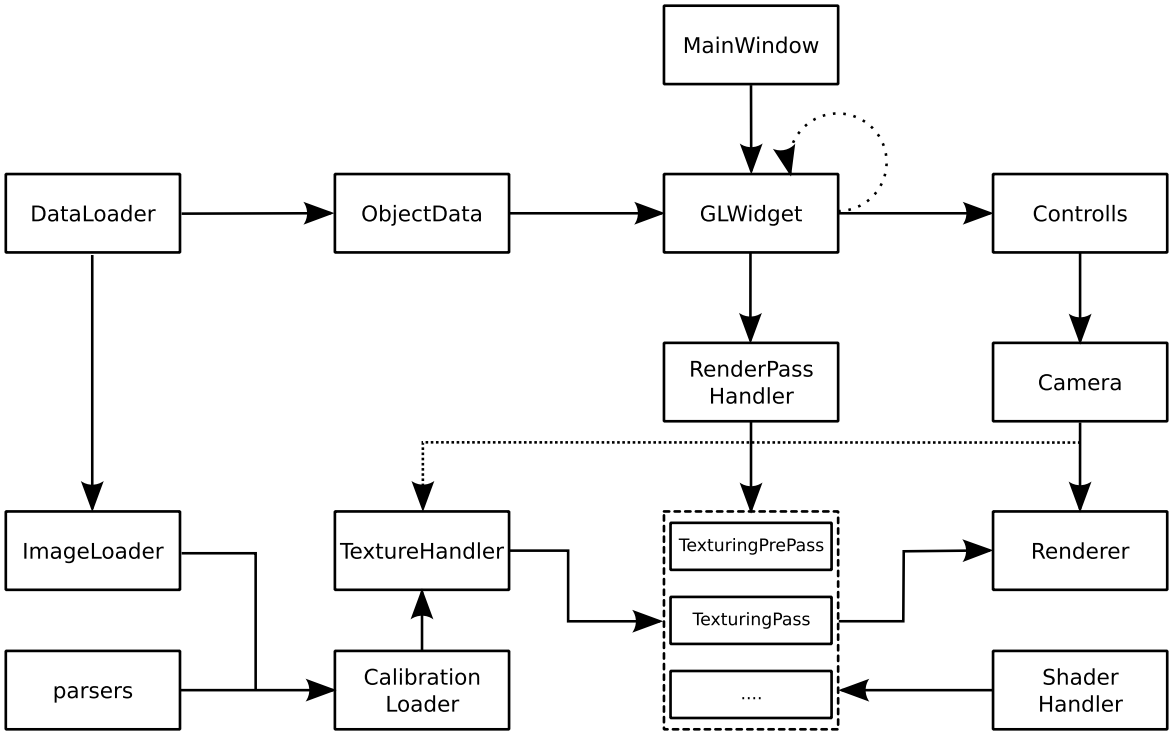
3.1 Použité technologie a knihovny

Aplikace je implementována v ANSI C++11 a je možné ji přeložit v linuxových operačních systémech i v MS Windows. Základ aplikace je postaven na multiplatformní knihovně Qt 4.8. Qt je jedna z nejrozšířenějších knihoven pro tvorbu aplikací s grafickým rozhraním (GUI), ale poskytuje i další podpůrné moduly, které s grafickým rozhraním přímo nesouvisí. Kromě tvorby GUI je Qt využito pro načítání vstupů z periférií a pro správu vláken. Pro efektivní zobrazování 3D grafiky je využito OpenGL, které poskytuje API pro rendering a výpočty na grafickém procesoru (GPU). Společně s OpenGL je použit jazyk GLSL, který slouží k tvorbě tzv. shaderů. To jsou samostatné programy, které se spouští na grafické kartě a řídí jednotlivé části programovatelného zobrazovacího řetězce. Pro běh aplikace je nutná grafická karta podporující OpenGL 4.0, což je jakákoliv novější grafická karta. Část programu využívá OpenGL 4.3, ale pro základní funkčnost aplikace není tato verze vyžadována.

Společně s OpenGL je využita knihovna GLEW 1.9, která slouží k načítání OpenGL rozšíření. S OpenGL je také úzce spjatá knihovna GLM, která poskytuje matematické operace běžně používané v počítačové grafice. Knihovna poskytuje základní datové struktury jako jsou vektory a matice a různé algebraické operace s těmito strukturami. Další použité knihovny jsou libjpeg 8 pro efektivní načítání obrázku ve formátu JPG a Assimp 3 pro načítání 3D modelů ve velkém množství formátů. Všechny využité knihovny jsou dostupné pod otevřenou licencí. glm

3.2 Architektura aplikace

Architektura aplikace zobrazující 3D grafiku v reálném čase se liší od běžných desktopových aplikací. Základem je efektivně využít zobrazovací řetězec, který OpenGL poskytuje, a zajistit distribuci dat mezi aplikací a grafickou kartou tak, aby byly omezeny přenosy dat mezi CPU a GPU, které jsou velmi drahé. Zároveň je důležité, aby všechny zobrazovací průchody měly přístup k aktuálním datům potřebným pro rendering. Samotné OpenGL je možné popsat jako data-driven architekturu, nicméně aplikace využívá tradičního objektového návrhu. Zjednodušený návrh architektury aplikace je zobrazený na obrázku 3.1.



Obrázek 3.1: Diagram ilustrující komunikaci základních komponent v aplikaci.

Základem aplikace je třída `MainWindow`, která se stará o všechny komponenty GUI. Obsahuje aplikační menu a stará se o otevírání dialogových oken s nastavením programu. Jako centrální widget obsahuje třídu `GLWidget`, která má na starost zobrazování dat vykreslených pomocí OpenGL. Třída obsahuje metodu `paintGL`, která slouží jako hlavní smyčka, ve které se překresluje scéna. Zároveň se také třída stará o vytvoření nové scény a inicializuje načtení vstupních dat.

3.3 Načítání vstupních dat

Aplikace na vstupu očekává 3D model a kalibrované fotografie. O načítání 3D modelu se stará třída `DataLoader` za pomoci knihovny Assimp. O načtení kalibrovaných fotografií se stará třída `CalibrationLoader`. Ta na základě zvoleného formátu vybírá parser pro načtení konfiguračních souborů s daty a stará se o přiřazení správných fotografií k načteným kamerám. Aplikace v současnosti podporuje dva formáty kalibrovaných dat:

- Bundler - výstupní soubor *.out z rekonstrukčního programu Bundler.
- REALVIZ Ascii Camera - soubor *.rz3 s kalibračními daty a textový soubor, který k datům z rz3 souboru přiřazuje názvy fotografií.

Vstupní fotografie jsou očekávány v nejběžnějším formátu JPG. Tento formát je však nevhodný pro textury, protože je komprimovaný, jeho načítání je netriviální a je příliš po-

malé. To by ani nebyl takový problém, pokud bychom fotografie načítali pouze jednou při spuštění aplikace. Problém je v tom, že fotografie mohou mít velké rozlišení a může jich být velké množství. Během testování byly použity scény s několika stovkami fotografií v rozlišení 4094×4096px. Takové množství fotografií v nekomprimovaném formátu zabere až desítky GB paměti. Nemůžeme proto očekávat, že je možné nahrát všechny fotografie do RAM a bude je nutné za běhu načítat z pevného disku. Z toho důvodu jsou při prvním načtení scény všechny fotografie načteny z původních JPG souborů a na disk se uloží každá fotografie také v nekomprimovaném formátu RAW. Při každém dalším spuštění se používají již nekomprimovaná data. Toto má nevýhodou v tom, že je nutné mít na disku dostatečné množství místa pro nekomprimovaná data. Nicméně pro běh aplikace v reálném čase je toto nezbytné.

Bohužel ani vytvoření RAW souborů není dostatečné pro načítání fotografií v reálném čase. Fotografie o rozměrech 4094×4096px má v nekomprimovaném formátu velikost přibližně 45MB. Moderní HDD jsou schopné číst data rychlostí přibližně 150MB/s, to znamená, že načtení fotografie z HDD bute trvat minimálně 0.3 sec. I když budeme uvažovat moderní SSD disky s rychlostí čtení přes 500MB/s, bude trvat načtení fotografie do RAM téměř 90ms. To je samozřejmě naprosto nepoužitelné pro zobrazování v reálném čase, kdy při 30 fps je nutné vyrenderovat celý snímek za 33ms.

Jako řešení se nabízí možnost předem načítat omezené množství fotografií, které se vejde do RAM, aby fotografie byla vždy již načtená ve chvíli, kdy je jí třeba použít. Toto jsme implementovali za pomoci kd-stromu, kdy se načítaly fotografie nejbližší k aktuální pozici. Toto řešení se ukázalo velmi neefektivní, protože nebylo možné načítat dostatečné množství fotografií pro pokrytí všech možných směrů pohybu kamery. Ve výsledku byl procesor vytížený neustálým načítáním fotografií a aplikace byla nepoužitelná. Proto jsme zvolili jiný přístup. Pro každou fotografii jsou vytvořeny náhledy o velikosti šířce 512px. Tyto náhledy je již snadno možné nahrát všechny do RAM. Při texturování se poté využívají náhledy a originální fotografie je načtena pouze v případě, kdy je opravdu použita. To vede ke snížení kvality textur při změně pohledu virtuální kamery, ale pokud se pohled virtuální kamery příliš nemění, textury v plném rozlišení se načtou poměrně rychle (konkrétní hodnoty jsou uvedeny v kapitole testování v sekci TODO).

-

promítnutí bodů do fotek, konvexní obálka, úprava úhlu

Kapitola 4

Testování

Kapitola 5

Závěr

Literatura

[AB91] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[AMK10] Ehsan Aganj, Pascal Monasse, and Renaud Keriven. Multi-view texturing of imprecise mesh. In *Computer Vision – ACCV 2009*, volume 5995 of *Lecture Notes in Computer Science*, pages 468–476. Springer Berlin Heidelberg, 2010.

[And79] A. M. Andrew. Another efficient algorithm for convex hulls in two dimensions. *Inf. Process. Lett.*, 9(5):216–219, 1979.

[APK08] C. Allene, J.-P. Pons, and R. Keriven. Seamless image-based texture atlases using multi-band blending. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.

[BMR01] Fausto Bernardini, Ioana M. Martin, and Holly Rushmeier. High-quality texture reconstruction from multiple scans. *IEEE Transactions on Visualization and Computer Graphics*, 7:318–332, 2001.

[CCCS08] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno. Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3d models. *Comput. Graph.*, 32(4):464–473, August 2008.

[DTM96] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. SIGGRAPH '96, pages 11–20, New York, NY, USA, 1996. ACM.

[EDDM+08] Martin Eisemann, Bert De Decker, Marcus Magnor, Philippe Bekaert, Edilson de Aguiar, Naveed Ahmed, Christian Theobalt, and Anita Sellent. Floating textures. *Computer Graphics Forum (Proc. of Eurographics)*, 27(2):409–418, April 2008.

[HZ04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[LI07] V. Lempitsky and D. Ivanov. Seamless mosaicing of image-based texture maps. In *Computer Vision and Pattern Recognition*, pages 1–6, 2007.

[SD02] Marc Stamminger and George Drettakis. Perspective shadow maps. *ACM Trans. Graph.*, 21(3):557–562, July 2002.

[Sze10] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010.

[TM09] Takeshi TAKAI and Takashi MATSUYAMA. Harmonized texture mapping. *The Journal of The Institute of Image Information and Television Engineers*, 63(4):488–499, apr 2009.

[ŽBSF04] J. Žára, B. Beneš, J. Sochor, and P. Felkel. *Moderní počítačová grafika*. Computer Press, 2004.

Příloha A

Seznam použitých zkratek

2D Two-Dimensional

ABN Abstract Boolean Networks

ASIC Application-Specific Integrated Circuit

⋮

Příloha B

Instalační a uživatelská příručka

B.0.1 Překlad aplikace

B.0.2 Používání aplikace

Příloha C

Obsah přiloženého CD