



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ**

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

**VÝPOČET MAPY DISPARITY ZE STEREO OBRAZU**

DISPARITY MAP ESTIMATION FROM STEREO IMAGE

**DIPLOMOVÁ PRÁCE**

MASTER'S THESIS

**AUTOR PRÁCE**

AUTHOR

**Bc. ROMAN TÁBI**

**VEDOUcí PRÁCE**

SUPERVISOR

**Ing. MICHAL ŠPANĚL, Ph.D.**

BRNO 2017

**Vysoké učení technické v Brně - Fakulta informačních technologií**

Ústav počítačové grafiky a multimédií

Akademický rok 2016/2017

**Zadání diplomové práce**

Řešitel: **Tábi Roman, Bc.**

Obor: Počítačová grafika a multimédia

Téma: **Výpočet mapy disparity ze stereo obrazu**  
**Disparity Map Estimation from Stereo Image**

Kategorie: Počítačová grafika

**Pokyny:**

1. Prostudujte problematiku stereo vidění a 3D rekonstrukce ze stereo obrazu. Zorientujte se v současných metodách výpočtu tzv. disparity map ze stereo obrazu.
2. Seznamte se s principy konvolučních neuronových sítí (CNN) a možnostmi jejich využití pro výpočet mapy disparity.
3. Vyberte vhodné metody a navrhnete řešení pro výpočet disparity mapy s využitím konvolučních sítí.
4. Experimentujte s Vaší implementací a případně navrhnete vlastní modifikace metod.
5. Porovnejte dosažené výsledky a diskutujte možnosti budoucího vývoje.
6. Vytvořte stručný plakát prezentující Vaši diplomovou práci, její cíle a výsledky.

**Literatura:**

- Dle pokynů vedoucího

Při obhajobě semestrální části projektu je požadováno:

- Splnění prvních tří bodů zadání.

Podrobné závazné pokyny pro vypracování diplomové práce naleznete na adrese

<http://www.fit.vutbr.cz/info/szz/>

Technická zpráva diplomové práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap, které byly vyřešeny v rámci dřívějších projektů (30 až 40% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Španěl Michal, Ing., Ph.D., UPGM FIT VUT**

Datum zadání: 1. listopadu 2016

Datum odevzdání: 24. května 2017

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
Fakulta informačních technologií  
Ústav počítačové grafiky a multimédií  
602 00 Brno, Božetěchova 2



---

doc. Dr. Ing. Jan Černocký  
vedoucí ústavu

## Abstrakt

Diplomová práca sa zameriava na výpočet mapy disparity s použitím konvolučnej neurónovej siete. Preberá problematiku použitia konvolučných neurónových sietí pre porovnanie obrazov a výpočet disparity zo stereo obrazu ako aj existujúce prístupy pre riešenie zvoleného problému. Navrhuje a implementuje systém pozostávajúci z konvolučnej neurónovej siete pre odhad podobnosti dvoch výrezov obrazu, a metód pre filtráciu a vyhladenie výslednej mapy disparity. Experimenty a výsledky ukázali, že najkvalitnejšie disparitné mapy generuje riešenie, kde neurónová sieť porovnáva výrezy s rozmermi  $9 \times 9$  pixlov v spojení s algoritmom pre agregáciu a korekciu párovacej ceny a bilaterálnym filtrom.

## Abstract

The master thesis focuses on disparity map estimation using convolutional neural network. It discusses the problem of using convolutional neural networks for image comparison and disparity computation from stereo image as well as existing approaches of solutions for given problem. It also proposes and implements system that consists of convolutional neural network that measures the similarity between two image patches, and filtering and smoothing methods to improve the result disparity map. Experiments and results show, that the most quality disparity maps are computed using CNN on input patches with the size of  $9 \times 9$  pixels combined with matching cost aggregation and correction algorithm and bilateral filter.

## Kľúčové slová

konvolučné neurónové siete, stereo obraz, disparita, mapa disparity, caffe

## Keywords

convolutional neural networks, stereo image, disparity, disparity map, caffe

## Citácia

TÁBI, Roman. *Výpočet mapy disparity ze stereo obrazu*. Brno, 2017. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Španěl Michal.

# Výpočet mapy disparity ze stereo obrazu

## Prehlásenie

Prehlasujem, že som túto diplomovú prácu vypracoval samostatne pod vedením Ing. Michala Španěla, Ph.D. Uviedol som všetky literárne pramene a publikácie, z ktorých som čerpal.

.....  
Roman Tábi  
21. mája 2017

## Podakovanie

Podakovanie patrí predovšetkým vedúcemu diplomovej práce p. Ing. Michalovi Španělovi, Ph.D. za jeho ochotu, čas, pomoc a cenné rady pri vypracovaní tejto práce. Tak isto ďakujem za pomoc p. Ing. Michalovi Hradišovi, Ph.D, a svojej rodine a priateľom za podporu. Ďalej veľmi oceňujem prístup k výpočtovým a úložným zariadeniam, ktoré vlastní strany a projekty prispievajúce k Národnej Gridovej Infraštruktúre MetaCentrum, a ktoré sú poskytované v rámci programu "Projekty Velkého Výskumu, Vývoja a Inovatívnych Infraštruktúr".



# Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Stereo videnie</b>	<b>3</b>
2.1	Epipolárna geometria . . . . .	3
2.2	Rektifikácia obrazu . . . . .	5
<b>3</b>	<b>Disparita a disparitná mapa</b>	<b>8</b>
3.1	Metriky kvality disparitných máp . . . . .	8
<b>4</b>	<b>Konvolučné neurónové siete</b>	<b>11</b>
4.1	Neurónové siete . . . . .	11
4.2	Konvolučné neurónové siete CNN . . . . .	13
4.3	Frameworky pre trénovanie CNN . . . . .	16
<b>5</b>	<b>Súčasný stav</b>	<b>18</b>
5.1	Metóda PMSC . . . . .	19
5.2	Metóda MeshStereoExt . . . . .	20
5.3	Metóda MC-CNN-acrt . . . . .	21
<b>6</b>	<b>Návrh riešenia výpočtu mapy disparity zo stereo obrazu</b>	<b>23</b>
6.1	Konvolučná neurónová sieť pre porovnanie dvojice obrázkov . . . . .	23
6.2	Stereo metóda . . . . .	24
<b>7</b>	<b>Implementácia</b>	<b>29</b>
<b>8</b>	<b>Experimenty a výsledky</b>	<b>31</b>
8.1	Konvolučná neurónová sieť pre porovnanie dvojice obrázkov . . . . .	31
8.2	Stereo metóda . . . . .	34
<b>9</b>	<b>Záver</b>	<b>43</b>
	<b>Literatúra</b>	<b>44</b>
<b>A</b>	<b>Plagát</b>	<b>47</b>

# Kapitola 1

## Úvod

Rekonštrukcia 3D scény je kľúčovou súčasťou mnohých aplikácií, napr. robotických systémov alebo autonómnych áut. Riešením zvyčajne sú 3D senzory ako LIDAR, ale alternatívou najmä vďaka nižšej cene môže byť použitie kamier. Aj napriek desiatkam rokov výskumu a vývoja je odhad disparity zo stereo obrazu stále nevyriešený problém [16]. V tejto práci je problematika odhadu disparity zo stereo obrazu rozoberaná a riešená s použitím konvolučných neurónových sietí.

Cielom tejto práce je navrhnúť a natrénovať konvolučnú neurónovú sieť vhodnú pre použitie pri výpočte disparity a vytvoriť aplikáciu pre výpočet mapy disparity zo stereo obrazu s použitím natrénovanej siete.

Riešenie vychádza z postupov a princípov popísaných v [28] a [26]. Použitá je konvolučná neurónová sieť so siamskou architektúrou. Neurónová sieť je použitá pre výpočet párovacej ceny výrezov obrazov, ktorá je základom pre výpočet disparitnej mapy. V ďalších krokoch popisovaného riešenia prebieha zhlukovanie a korekcia ceny ako aj filtrácia výslednej disparitnej mapy.

Experimenty sú prevádzané na sieťach so vstupnými obrazmi s rôznymi rozmermi, a výsledné mapy sú porovnávané so vzorovými. Vylepšovaním výstupu siete a samotných disparitných máp je dosiahnutých menej ako 30% chybných určených pixlov voči vzorovým mapám.

Technická správa je rozdelená na kapitoly, z ktorých každá popisuje špecifickú časť procesu vývoja riešenia problému. V kapitolách 2, 3 a 4 sú popisované dôležité princípy a komponenty používané v tejto práci. Súhrn metód, používaných pri riešení výpočtu mapy disparity, je obsahom kapitoly 5. Popis vytvoreného riešenia v tejto práci je v kapitolách 6 a 7, a experimenty a testovanie sú predmetom kapitoly 8.

# Kapitola 2

## Stereo videnie

Emulácia stereo videnia ľudského vizuálneho systému počítačmi je prevádzaná hľadáním korešpondencií medzi bodmi v dvoch obrazoch. S takýmito bodmi a známou vzdialenosťou medzi kamerami môžeme vypočítať 3D pozície týchto bodov. Hoci takéto hľadanie bodov môže byť výpočtovo veľmi náročné, môžeme využiť znalosti o geometrii scény a výrazne obmedziť prehľadávaný priestor. Počítačové stereo zobrazovanie zahŕňa niekoľko krokov, ktoré musia byť prevedené pre rekonštrukciu 3D priestoru [6]:

- Matematicky odstrániť skreslenie šošoviek.
- Opraviť uhol a vzdialenosti kamier. Tento proces sa nazýva *rektifikácia*.
- Nájsť korešpondujúce znaky v ľavom a pravom obraze. Výstupom tohoto kroku je *disparitná mapa*.
- Pri známom geometrickom postavení kamier je možné transformovať disparitnú mapu na hĺbkovú mapu.

### 2.1 Epipolárna geometria

Epipolárna geometria je vnútorná projekčná geometria medzi dvomi pohľadmi. Je nezávislá na štruktúre scény, a závisí len od vnútorných parametrov kamier a ich relatívnym pozíciám.

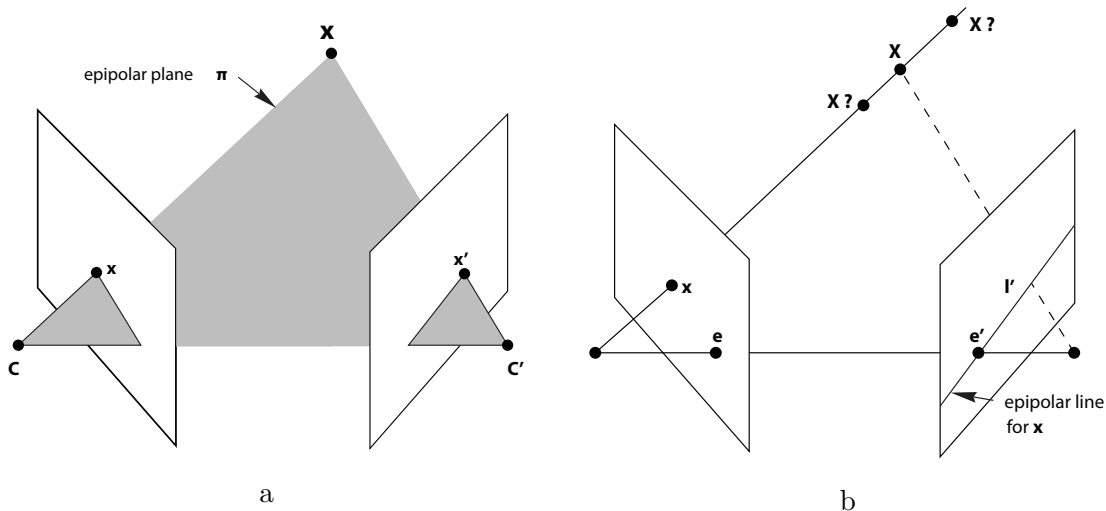
Fundamentálna matica  $F$  popisuje túto vnútornú geometriu. Matica má rozmer  $3 \times 3$  a hodnotu 2. Ak je bod v  $X$  trojdimenzionálnom priestore premietaný ako  $x$  v pohľade prvej kamery, a  $x'$  v pohľade druhej kamery, potom body videných obrazov spĺňajú vzťah

$$x'^T F x = 0 \tag{2.1}$$

Epipolárna geometria medzi dvomi pohľadmi je v podstate geometriou prieniku obrazov na priemetniach, ktorej základňou je priamka spájajúca stredy kamier. Táto geometria sa využíva pri hľadaní korešpondujúcich bodov pri stereo videní.

Daný je bod  $X$  v 3D priestore premietaný do dvoch obrazov, ako  $x$  v prvom a  $x'$  v druhom. Aký je vzťah medzi korešpondujúcimi bodmi  $x$  a  $x'$ ? Na obr. 2.1 je vidieť, že premietnuté body  $x$  a  $x'$ , bod v priestore  $X$ , a stredy kamier  $C$  a  $C'$  sú v jednej rovine označenej  $\pi$ . Je zrejmé, že spätne premietnuté lúče z  $x$  a  $x'$  pretínajú bod  $X$ , sú v jednej rovine a ležia v  $\pi$ . Pri hľadaní podobností je táto rovina  $\pi$  najdôležitejšou veličinou.

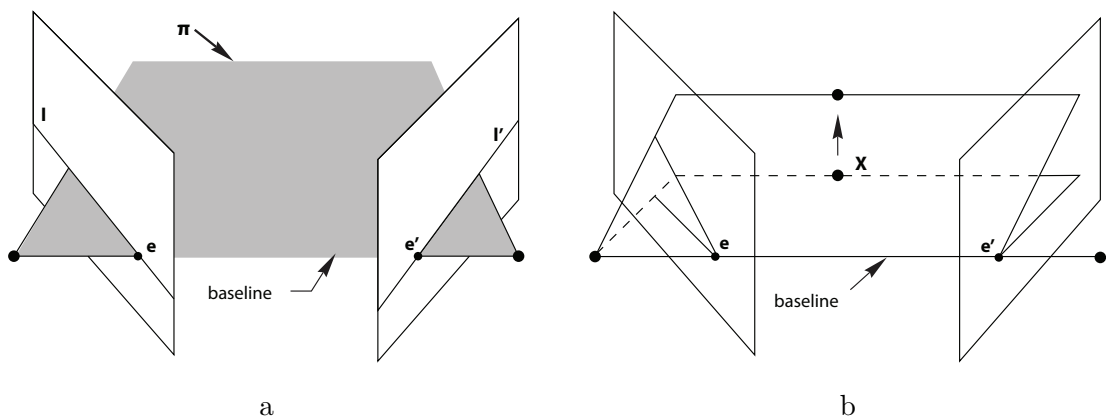
Pri hľadaní korešpondencií poznáme len bod  $x$  a hľadáme bod  $x'$ . Rovina  $\pi$  je určená priamkou spájajúcou stredy kamier a lúčom definovaným bodom  $x$ . Z vyššie uvedeného



Obr. 2.1: Geometria korešpondujúcich bodov. (a) Kamery sú označené ich stredmi  $C$  a  $C'$  a ich priemetňami. Stredy kamier, priestorový bod  $X$  a jeho obrazy  $x$  a  $x'$  ležia v jednej rovine  $\pi$ . (b) Priemietnutý bod  $x$  a stred kamery  $C$  definujú spätne priemietnutý lúč. Tento lúč je zobrazený do priamky  $l'$  v obraze druhej kamery. 3D bod  $X$ , ktorý je priemietnutý ako  $x$  musí ležať na tomto lúči, takže obraz bodu  $X$  v druhom obraze musí ležať na priamke  $l'$  [10].

vyplýva, že lúč prislúchajúci neznámemu bodu  $x'$  leží v  $\pi$ , a teda bod  $x'$  leží na priamke  $l'$  danej prienikom roviny  $\pi$  a priemetne druhej kamery. Táto priamka  $l'$  je obrazom lúča definovaným bodom  $x$ , a nazýva sa *epipolárna čiara* prislúchajúca bodu  $x$ . Algoritmy hľadania korešpondencií v stereo obrazoch využívajú tieto princípy tak, že pri hľadaní bodu prislúchajúceho bodu  $x$  nepotrebnú prechádzať celý obraz druhej kamery, ale postačuje spracovávať body ležiace na priamke  $l'$ .

Geometrické entity zahrnuté v epipolárnej geometrii sú ilustrované na obr. 2.2.

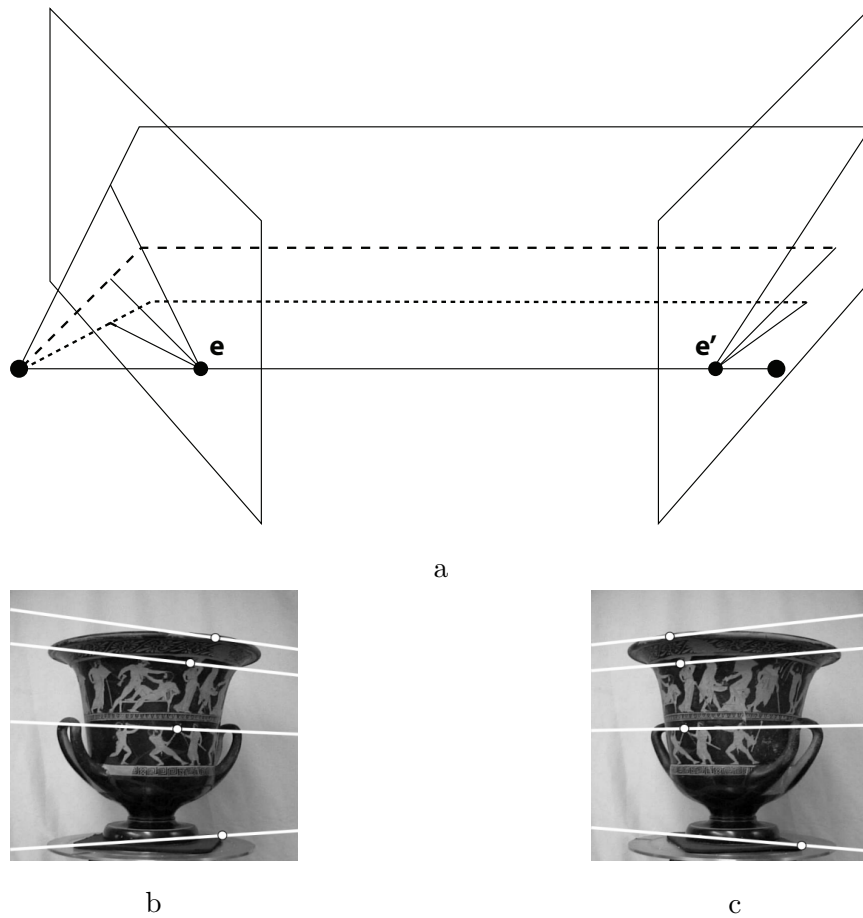


Obr. 2.2: Epipolárna geometria. (a) Priamka spájajúca stredy kamier pretína obidve priemetne v epipóloch  $e$  a  $e'$ . Každá rovina  $\pi$  obsahujúca túto priamku je epipolárna rovina a pretína priemetne v odpovedajúcich epipolárnych čiarach  $l$  a  $l'$ . (b) Zmenou pozície 3D bodu  $X$  epipolárne roviny rotujú okolo základne. Všetky epipolárne čiary sa pretínajú v epipóle [10].

Zavedená terminológia týchto entít je nasledovná:

- Epipól je bod prieniku základne (priamky spájajúcej stredy kamier) s priemetňou.
- Epipolárna rovina je rovina obsahujúca základňu.
- Epipolárna čiara je prienik epipolárnej roviny s priemetňou. Všetky epipolárne čiary sa pretínajú v epipóle. Epipolárna rovina prechádza ľavou a pravou priemetňou v epipolárnych čiarach a definuje korešpondencie medzi čiarami.

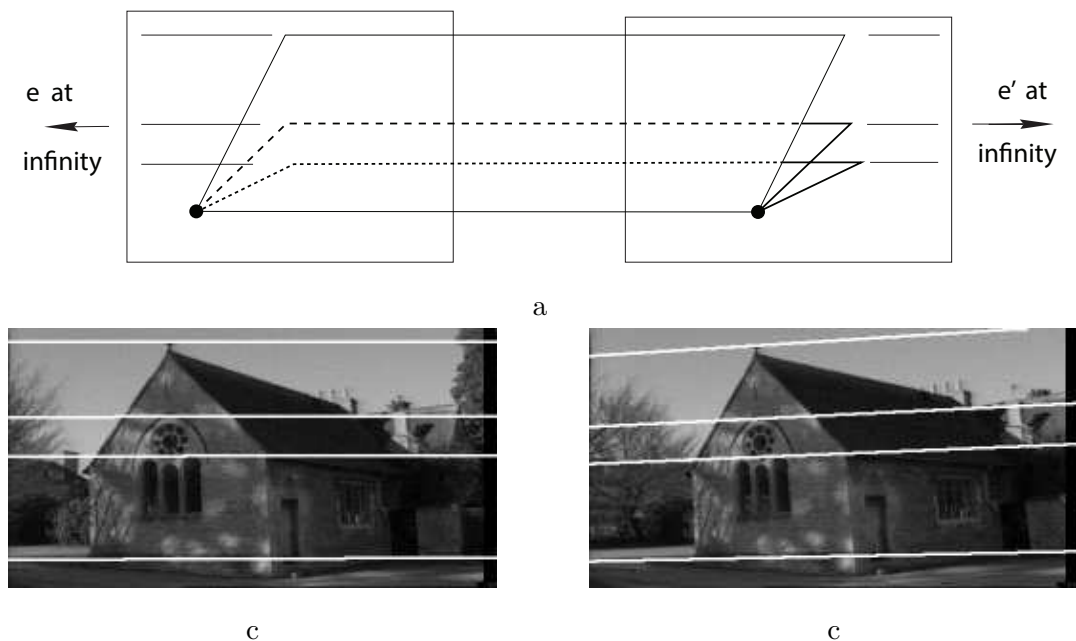
Ukážky epipolárnej geometrie sú na obr. 2.3 a obr. 2.4.



Obr. 2.3: Konvergujúce kamery. (a) Epipolárna geometria pre konvergujúce kamery. (b) a (c) Pár obrazov so zobrazenými korešpondujúcimi bodmi a ich epipolárnymi čiarami. Transformačný pohyb medzi pohľadmi je posun a rotácia. V každom obraze môže byť odhadnutý smer druhej kamery prienikom množiny epipolárnych rovín epipolárnych čiar. V tomto prípade ležia obidva epipóly mimo obrazov [10].

## 2.2 Rektifikácia obrazu

Hľadanie najlepšej zhody pozdĺž epipolárnej čiary môže byť veľmi efektívne v prípade, že epipolárne čiary sú rovnobežné s horizontálnou osou obrazu a kolineárne (odpovedajúce



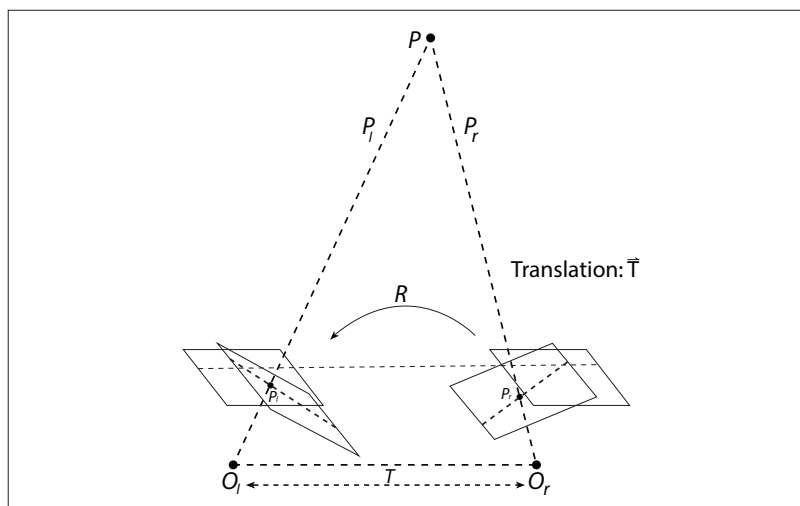
Obr. 2.4: Paralelný pohyb voči priemetni. V tomto prípade pohybu, kde translácia je paralelná na priemetňu a rotačná os je kolmá na priemetňu, prienik základne s priemetňou je v nekonečne. Z toho vyplýva, že epipóly sú v nekonečne a epipolárne čiary sú rovnobežné. (a) Epipolárna geometria pre pohyb paralelný voči priemetni. (b) a (c) Pár obrazov, kde pohyb medzi obrazmi je (približne) translácia paralelná na os  $x$ , bez rotácie. Zobrazené sú epipolárne čiary. Korešpondujúce body ležia na odpovedajúcich epipolárnych čiarach [10].

prehľadávané čiary v oboch obrazoch). Pre takúto konfiguráciu, hľadanie bodu prislúchajúcejmu bodu  $(x, y)$  prebieha len na pixloch  $(*, y)$ , teda v horizontálnom smere. Takáto konfigurácia sa nazýva jednoduchý alebo štandardný stereo systém. Obrazy zachytené dvomi rôznymi kamerami potom môžu byť transformované tak, aby obsahovali epipolárne čiary rovnobežné s horizontálnou osou procesom, ktorý sa nazýva rektifikácia [25].

Disparita zo stereo obrazu sa najjednoduchšie vypočíta, ak sú priemetne kamier dokonale zarovnané. Nanešťastie v praxi sú dokonale zarovnané kamery veľmi zriedkavo, keďže kamery takmer nikdy nemajú presne rovinné, riadkovo zarovnané priemetne. Obr. 2.5 zobrazuje zámer rektifikácie: cieľ je znovu premietnuť obraz na priemetňach oboch kamier do jednej roviny tak, aby boli riadkovo zarovnané.

Pre zarovnanie priemetní sú pre každú kameru potrebné parametre, a to vektor skreslenia, rotačná matica, a rektifikovaná a nerektifikovaná matica kamery. Existuje niekoľko algoritmov pre výpočet týchto parametrov, napr. Hartleyho algoritmus [9].

Po prevedení transformácií obrazov sa ich epipóly nachádzajú v nekonečne. Po zarovnaní riadkov oboch obrazov (epipolárne čiary rovnobežné s horizontálnou osou) môže hľadanie korešpondenčných bodov stereo algoritmom prebiehať v jednom riadku pre daný bod [6].



Obr. 2.5: Cieľom je matematicky zarovnať priemetne oboch kamier do jedinej roviny [6].

## Kapitola 3

# Disparita a disparitná mapa

Človek vníma svet v troch dimenziách aj napriek tomu, že vstupy do nášho vizuálneho systému, obrazy premietané na sietnicu oka, sú dvojdimenzionálne. Tento jav je dosiahnutý schopnosťou vizuálneho systému odhadovať tretiu dimenziu, hĺbku, pomocou rôznych vizuálnych znakov zo sietnicových obrazov. Jedným z týchto znakov je stereoskopická disparita, zmena pozície medzi dvomi sietnicovými projekciami bodu v priestore (obr. 3.1). Táto zmena je spôsobená rôznymi pozíciami očí v horizontálnom smere.

Pre odhad stereoskopickej disparity musí vizuálny systém vyriešiť korešpondenčný problém: musí vedieť rozhodnúť, ktoré časti dvoch sietnicových obrazov prislúchajú rovnakému objektu reálneho sveta. Historicky bolo zavedené, že vizuálny systém rieši tento problém párovaním významných črt oboch obrazov. Napriek tomu však vzniklo množstvo stereo algoritmov, ktoré popisujú riešenie korešpondenčného problému rôznymi spôsobmi ([17], [19], [21], ...) [20].

Disparita  $d$  v počítačovom stereo videní reprezentuje zmenu v horizontálnej polohe objektu v ľavom a v pravom obraze – objekt na pozícii  $(x, y)$  v ľavom obraze sa nachádza na pozícii  $(x - d, y)$  v pravom obraze. Je preto potrebné, aby boli spracovávané obrazy rektifikované, epipolárne čiary rovnobežné s horizontálnou osou obrazu, aby bolo možné prevádzať výpočet disparity v horizontálnom smere. Ak poznáme disparitu objektu, jeho hĺbku  $z$  môžeme vypočítať pomocou vzťahu:

$$z = \frac{fB}{d} \quad (3.1)$$

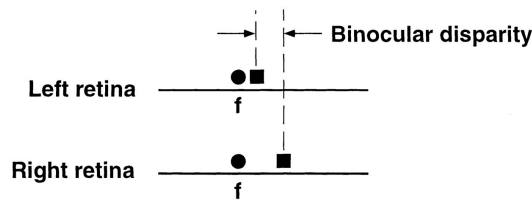
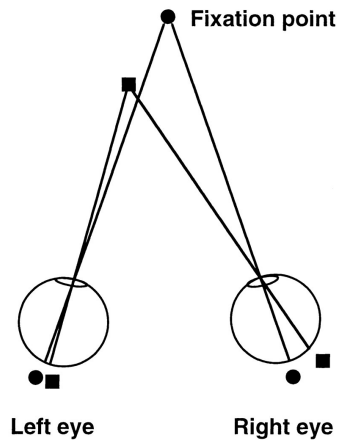
kde  $f$  je ohnisková vzdialenosť kamery a  $B$  je vzdialenosť medzi kamerami.

Disparitnú mapu dvojice obrazov reprezentuje ľavý, resp. pravý obraz, v ktorom každý pixel obsahuje hodnotu disparity pre daný pixel v pravom, resp. ľavom, obraze. Obr. 3.2 zobrazuje vstupy a výstup algoritmu pre výpočet mapy disparity [28].

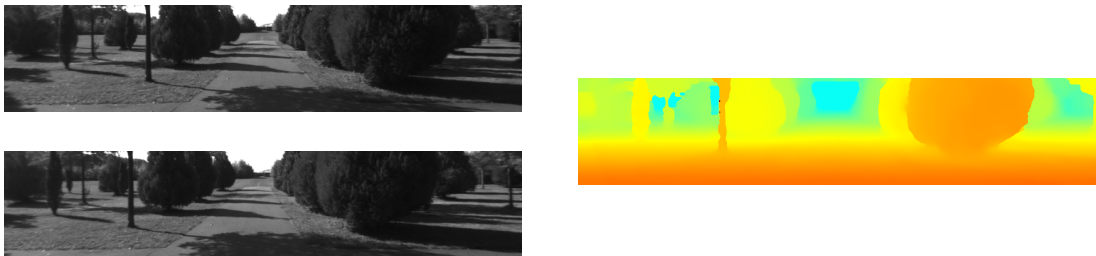
### 3.1 Metriky kvality disparitných máp

Pre ohodnotenie výstupov stereo algoritmov alebo efektov rôznych parametrov na výstup, sú potrebné postupy pre odhad kvality spočítaných korešpondencií. Dva obecné postupy sú výpočet štatistiky chýb voči určitým vzorovým dátam (obr. 3.3), a ohodnotenie syntetizovaných obrazov deformovaním referenčných obrazov podľa vypočítaných hodnôt disparity.





Obr. 3.1: Geometria stereoskopickkej projekcie (hore) a definícia stereoskopickkej disparity (dole). Fixačný bod je pre mietaný do prislúchajúcich foveí a má nulovú disparitu podľa definície. Ostatné body nie sú premietané na príslušné pozície na sietniciach a majú nenulovú disparitu [20].



Obr. 3.2: Vstupom (vľavo) je pár obrazov z ľavej a pravej kamery. Obrazy sa líšia najmä v horizontálnych pozíciách objektov. Objekty bližšie ku kamere majú vyššiu hodnotu disparity ako vzdialenejšie objekty. Výstupom (vpravo) je mapa disparity, s teplejšími farbami reprezentujúcimi vyššiu hodnotu disparity [28].

Navrhnutý systém v tejto práci je testovaný a porovnávaný s inými metódami na internetovom portáli pre meranie kvality stereo algoritmov Middlebury Stereo Evaluation<sup>1</sup>. Software vypočítava nasledujúce miery kvality podľa vzorových dát:

1. *rms* – RMS (root-mean-squared) chyba medzi vypočítanou mapou disparity  $d_C(x, y)$  a vzorovou disparitnou mapou  $d_T(x, y)$ ,

<sup>1</sup>Middlebury Stereo Evaluation <http://vision.middlebury.edu/stereo/eval3/>



Obr. 3.3: Vstupný obraz z ľavej kamery a vzorová disparitná mapa (ground truth) prislúchajúca danému vstupu [23].

$$R = \left( \frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{\frac{1}{2}}, \quad (3.2)$$

kde  $N$  je celkový počet pixlov.

2.  $bad_t$  – Percento nesprávne spárovaných pixlov,

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_T(x,y)| > \delta_d), \quad (3.3)$$

kde  $\delta_d(eval\_bad\_thresh)$  je tolerancia chyby disparity,  $\delta_d \in \{0.5, 1.0, 2.0, 4.0\}$  [22].

3. Mediánová absolútna odchylka.

4. Celkový čas výpočtu, čas výpočtu normalizovaný počtom pixlov, čas výpočtu normalizovaný počtom úrovni disparity [23].

## Kapitola 4

# Konvolučné neurónové siete

Neurónové siete sú populárnym prostriedkom pre popis učenia. Tieto siete sú inšpirované neurónmi v ľudskom mozgu, ale v skutočnosti ich nesimulujú. Umelé neurónové siete typicky obsahujú výrazne menej ako  $10^{11}$  neurónov v ľudskom mozgu, a umelé neuróny, nazývané jednotky, sú jednoduchšie ako ich biologické náprotivky.

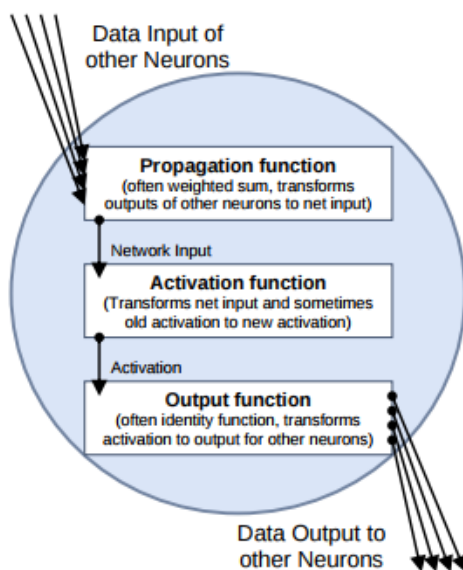
### 4.1 Neurónové siete

Umelé neurónové siete sú zaujímavou oblasťou pre štúdium kvôli viacerým dôvodom:

- Ako časť neurovedy, pre pochopenie skutočných neurónových systémov, vedci simulujú neurónové systémy jednoduchších živočíchov ako červy, ktoré smerujú k pochopeniu, ktoré aspekty neurónových systémov sú potrebné pre vysvetlenie správania týchto živočíchov.
- Niektorí vedci sa pokúšajú automatizovať nie len funkcionality inteligencie (ktorá je podstatou umelej inteligencie), ale taktiež mechanizmus mozgu. Jedna hypotéza hovorí, že jediným spôsobom, ako simulovať funkcionality mozgu je použitie mechanizmu mozgu. Táto hypotéza môže byť testovaná pokusmi vytvoriť inteligenciu s použitím mechanizmu mozgu, ale aj bez. Podľa iných systémov, napr. lietajúce zariadenia, ktoré využívajú rovnaké princípy, ale nie rovnaké mechanizmy, ako vtáky využívajú k lietaniam, môže napovedať, že táto hypotéza nemusí byť pravdivá. Napriek tomu je však zaujímave hypotézu testovať.
- Mozgom sa inšpirujú nové spôsoby výpočtov, ktoré sa líšia od bežne dostupných počítačov. Na rozdiel od terajších počítačov, ktoré majú niekoľko procesorov a veľké, ale v zásade pasívne, pamäte, mozog pozostáva z veľkého množstva asynchrónne distribuovaných procesov, ktoré bežia súčasne bez akéhokoľvek master kontroléra. Architektúra terajších počítačov teda nemusí byť jediným spôsobom pre výpočty.
- Neurónové siete poskytujú odlišnú mieru jednoduchosti pri trénovacej chybe, ako napr. rozhodovacie stromy. Viacvrstvové neurónové siete, ako aj rozhodovacie stromy, môžu reprezentovať funkciu na množine diskretných znakov. Avšak funkcie prislúchajúce jednoduchým neurónovým sieťam nemusia striktné prislúchať jednoduchým rozhodovacím stromom. Trénovanie neurónových sietí využíva odlišnú chybu ako učenie rozhodovacích stromov. Ktorá je lepšia v praxi, je otázka, ktorá môže byť testovaná na rôznych doménach [13].

Technická neurónová sieť pozostáva z jednoduchých výpočtových jednotiek, neurónov, a orientovaných, váhových prepojení medzi nimi. Váhy môžu byť reprezentované štvorcovou váhovou maticou, alebo váhovým vektorom s číslom riadka matice udávajúcim miesto, kde prepojenie začína, a číslom stĺpca matice udávajúcim, ktorý neurón je cieľovým. Dáta sú prevádzané medzi neurónmi pomocou prepojení s váhami, ktoré ich zosilňujú alebo tlmia.

Pre neurón  $j$  (obr. 4.1) väčšinou existuje množstvo neurónov s prepojením do  $j$ , takých neurónov, ktoré prevádzajú svoj výstup do  $j$ . Pre neurón  $j$  propagačná funkcia obdrží výstupy  $o_{i_1}, \dots, o_{i_n}$  iných neurónov  $i_1, i_2, \dots, i_n$  (ktoré sú prepojené s  $j$ ), a transformuje ich s ohľadom na váhy  $w_{i,j}$  na vstupy siete  $net_i$ , ktoré môžu byť ďalej spracované aktivačnou funkciou. Potom vstup siete  $net_i$  je výsledkom propagačnej funkcie.



Obr. 4.1: Spracovanie dát v neuróne [12].

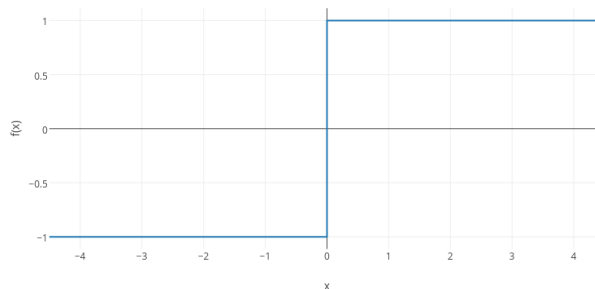
Podľa prírodného modelu každý neurón je stále do určitej miery aktívny alebo vybudený. Reakcie neurónov na vstupné hodnoty závisia od aktivačného stavu. Aktivačný stav udáva mieru neurónovho vybudenia a je často označovaný ako aktivácia. Aktivačná funkcia reaguje najcitlivejšie pri prahovej hodnote. Z biologického hľadiska prahová hodnota reprezentuje prah, pri ktorom neurón začína vysielat'. Pri určitom čase, aktivácia  $a_j$  neurónu  $j$  závisí od predchádzajúceho aktivačného stavu a externých vstupných hodnôt. Na rozdiel od iných premenných v neurónovej sieti, aktivačná funkcia je často globálne definovaná pre všetky neuróny, alebo pre množinu z nich a iba prahové hodnoty sa líšia pre každý neurón. Tieto prahové hodnoty sa môžu taktiež meniť počas tréningového procesu.

Najjednoduchšou aktivačnou funkciou je binárna prahovacia funkcia (obr. 4.2), ktorá môže generovať len 2 hodnoty. Používané sú aj sigmoida (obr. 4.3) alebo hyperbolický tangens (obr. 4.4).

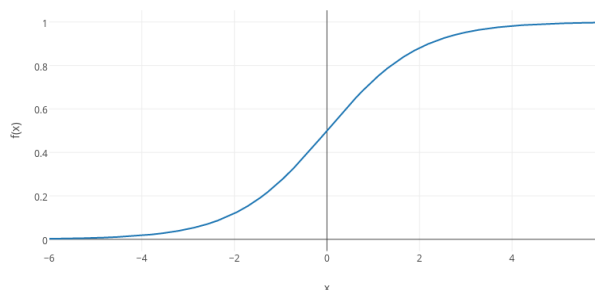
Výstupná funkcia neurónu vypočítava hodnoty, ktoré sú posúvané iným neurónom prepojeným s  $j$ . Vo všeobecnosti je výstupná funkcia tiež definovaná globálne. V mnohých prípadoch je táto funkcia identita, teda výstup aktivácie je priamo výstup neurónu.

Tréningový postup je algoritmus, ktorý je využitý k zmene a teda učeníu neurónovej siete, aby sieť produkovala očakávaný výstup pre daný vstup.

V tejto práci sa zaoberám doprednými neurónovými sieťami. V dopredných neurónových sieťach (obr. 4.5) sú neuróny rozložené do nasledujúcich vrstiev: vstupná vrstva,  $n$  skrytých



Obr. 4.2: Binárna prahovacia funkcia.



Obr. 4.3: Sigmoida.

výpočetných vrstiev a výstupná vrstva. V takýchto sieťach je každý neurón určitej vrstvy prepojený len s neurónmi nasledujúcej vrstvy (smerom k výstupnej vrstve). Vrstvy, v ktorých každý neurón  $i$  je prepojený so všetkými neurónmi nasledujúcej vrstvy sa nazývajú plne prepojené vrstvy.

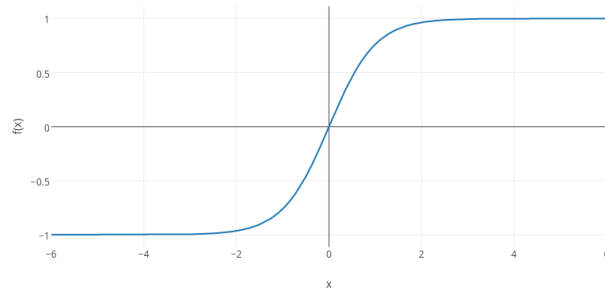
Množstvo neurónových sietí umožňuje zadávať vstupné dáta. Tieto dáta sú spracovávané sieťou, ktorá produkuje špecifický výstup. Neurónová sieť na obr. 4.5 má 2 vstupné neuróny a 2 výstupné neuróny, čo znamená, že sieť očakáva na vstupe 2 hodnoty a generuje 2 výstupné hodnoty [12].

## 4.2 Konvolučné neurónové siete CNN

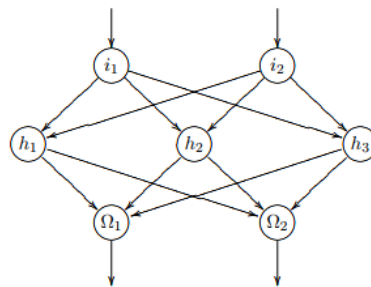
Konvolučné neurónové siete (CNN) inšpirované biologickými neurónovými sieťami sú používané na rozpoznávanie vzorov v obrazoch, napr. rozpoznanie tváre alebo rozpoznanie ručne písaného textu [18].

Vychádzajú z bežných neurónových sietí, sú tvorené neurónmi, ktoré obsahujú natrénované váhy a odchylky. Každý neurón obdrží vstupné dáta, prevádza skalárny súčin a voliteľne aplikuje nelienárnu funkciu (napr. ReLU). Produktom siete je skóre, ktoré určuje triedu vstupného obrazu siete.

Hlavným rozdielom medzi klasickými neurónovými sieťami a konvolučnými neurónovými sieťami je explicitný predpoklad, že vstupnými dátami sú obrazy 4.6.



Obr. 4.4: Hyperbolický tangens.

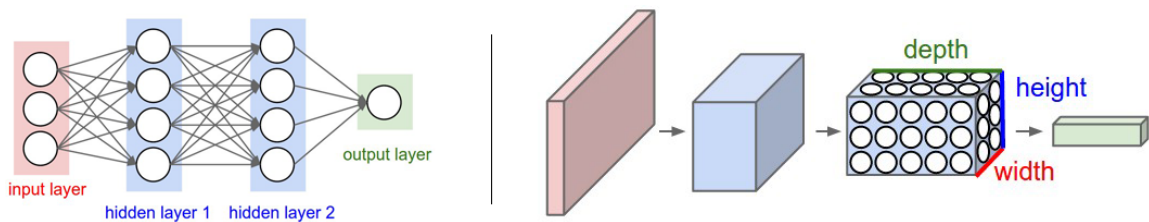


Obr. 4.5: Dopredná neurónová sieť s 3 vrstvami: vstupné neuróny  $i_x$ , skryté neuróny  $h_y$  a výstupné neuróny  $\Omega_z$  [12].

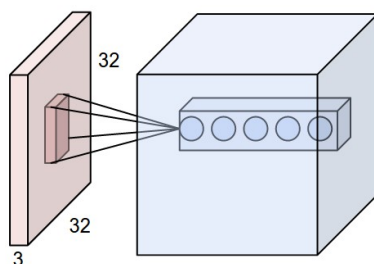
Jednoduchá konvolučná neurónová sieť je tvorená sekvenciou vrstiev, kde každá vrstva prevádza 3D vstup na obličný 3D výstup diferencovateľnou funkciou. Hlavné typy vrstiev, ktoré sa používajú pre vytvorenie architektúry siete sú konvolčná vrstva, deliaca vrstva, plne prepojená vrstva a aktivačná vrstva [2].

Konvolučná vrstva je hlavným stavebným prvkom konvolučnej neurónovej siete, ktorá prevádza väčšinu výpočtov. Parametre konvolučnej vrstvy pozostávajú z množiny filtrov schopných učiť sa. Každý filter je priestorovo malý (výškovo a šírko), ale rozšírený cez celú hĺbku vstupného objemu. Napr., typický filter prvej vrstvy konvolučnej neurónovej siete by mohol mať veľkosť  $5 \times 5 \times 3$  (5 pixlov šírku a výšku, a 3 z dôvodu hĺbky obrazu, farebných kanálov). Počas prechodu vpred sa posúva (konvuluje) každý filter po šírke a výške vstupu a vypočítavajú sa skalárne súčiny hodnôt filtra a vstupu vo všetkých pozíciách. Po posunutí filtra po celej ploche vstupu sa vytvorí 2-dimenzionálna aktivačná mapa, ktorá reprezentuje odozvy filtra v každej priestorovej polohe. Sieť sa naučí filtre, ktoré sa aktivujú, keď detekujú určitý typ vizuálneho vzoru ako hrana alebo škvrna určitej farby v prvej vrstve, alebo aj napr. kruhové vzory vo vyšších vrstvách. V konvolučnej vrstve je typicky množina takýchto filtrov (napr. 96 filtrov), každý z nich generuje vlastnú 2D aktivačnú mapu. Tieto mapy sú naskladané po hĺbkovej dimenzii a produkujú výstupný objekt (obr. 4.7) [2].

Kľúčovými technológiami konvolučných neurónových sietí sú veľkosti filtrov neurónov, zdieľanie váh, podzorkovanie podľa času alebo priestoru pre získanie špecifických znakov a redukciu veľkosti vstupných parametrov. Výhodyy CNN algoritmu sú implicitné učenie siete z tréningových dát, paralelné tréningovanie a redukcia zložitosti siete, aj jedinečné výhody pri rozpoznávaní reči a spracovaní obrazu [15].



Obr. 4.6: Vľavo: Klasická 3-vrstvová neurónová sieť. Vpravo: Konvolučná neurónová sieť usporadúva neuróny do 3 dimenzií (výška, šírka, hĺbka). Každá vrstva transformuje 3D vstup na 3D výstup [2].



Obr. 4.7: Príklad vstupného obrazu a neurónov v prvej konvolučnej vrstve. Každý neurón v konvolučnej vrstve je prepojený len k lokálnej oblasti vstupného obrazu priestorovo, ale do plnej hĺbky (všetky farebné kanály) [2].

Dôležitou súčasťou implementácie neurónových sietí je ich tréning. Tréning takýchto sietí môže byť zhrnuté do niekoľkých bodov:

1. Filtre a váhy sa inicializujú náhodnými hodnotami.
2. Sieť preberie obraz na vstupe a prevedie propagáciu v doprednom smere (konvólučie, ReLU a deliace operácie spolu s doprednou propagáciou v plne prepojených vrstvách) a vypočíta výsledné pravdepodobnosti pre všetky triedy. Pri prvej tréningovej iterácii sú výsledné pravdepodobnosti náhodné vzhľadom na náhodne vygenerované váhy.
3. Vypočíta sa celková odchylka na výstupnej vrstve.

$$c_o = \sum \frac{1}{2}(o_p - v_p)^2 \quad (4.1)$$

kde  $c_o$  je celková odchylka,  $o_p$  je očakávaná pravdepodobnosť a  $v_p$  je vypočítaná pravdepodobnosť.

4. Spätnou propagáciou chyby sú vypočítané gradienty chyby vo vzťahu ku všetkým váham v sieti a gradientným zostupom sú aktualizované všetky hodnoty filtrov pre minimalizáciu výstupnej odchylky. Váhy sú upravované podľa ich podielu na celkovej odchylke. Parametre ako veľkosti filtrov, počet filtrov alebo architektúra siete sa počas tréningu nemenia.
5. Opakujú sa kroky 2 - 4 so všetkými obrázkami v tréningovom datasete [11].

## 4.3 Frameworky pre tréovanie CNN

Pre tréovanie a používanie konvolučných neurónových sietí existuje množstvo rôznych knižníc a frameworkov. Vyvíjané sú v jazykoch ako napr. C/C++, Java, Python s podporou NumPy knižnice, ale aj Haskell, Scala, a i. Taktiež väčšina podporuje výpočty na GPU, obzvlášť na platforme CUDA od spoločnosti Nvidia, a tiež využívajú knižnicu cuDNN<sup>1</sup>. Medzi najrozšírenejšie frameworky pre tréovanie CNN patria nasledujúce:

### Caffe

Caffe je framework pre hlboké učenie vytvorený s dôrazom na rýchlosť a modularitu. Je vyvíjaný v Berkeley AI Research<sup>2</sup> laboratóriu a komunitou. Autorom projektu je Yangqing Jia<sup>3</sup>.

Framework umožňuje vývojárom vytvárať modely a optimalizovať jednoduchým prístupom konfiguračného rozhrania. Prepnutie nastavenia tréovania medzi CPU a GPU je možné jediným prepínačom.

Rozšíriteľný kód živí aktívny vývoj knižnice. V poslednom roku bol vetvený viac ako 1000 vývojármi, vďaka ktorým je aktívne udržiavaný v aktuálnom technologickom stave.

Rýchlosť tohto frameworku je ideálna pre experimenty vo vývoji ale aj v priemyselnom využití. Caffe dokáže spracovať viac ako 60 miliónov obrázkov za deň s grafickou kartou NVIDIA K40 GPU<sup>4</sup>. To odpovedá 1ms/obrázok pri rozhodovaní a 4ms/obrázok pri učení, pričom najnovšie verzie a hardware sú stále rýchlejšie [1].

### Caffe2

Caffe2 je framework pre hlboké učenie postavený na pôvodnom Caffe. Vyvíjaný v spolupráci so spoločnosťou Nvidia, je Caffe2 urýchľované najnovšími NVIDIA Pascal™ GPU a rozdeľuje sa medzi viacero GPU v rámci jedného uzlu. Tréovanie modelov môže teda zabráť hodiny namiesto dní [3].

### Torch

Torch je vedecký výpočtový framework s vysokou podporou algoritmov pre strojové učenie s dôrazom na prevádzanie výpočtov na GPU. Vďaka rýchlemu skriptovaciemu jazyku LuaJIT a C/CUDA implementácii je efektívny a jednoduchý na používanie.

Hlavnými znakmi frameworku Torch sú podpora výkonných N-dimenzionálnych polí, množstvo funkcií pre indexovanie, delenie, transponovanie, a pod. Taktiež funkcie pre lineárnu algebru, numerickú optimalizáciu, rýchla a efektívna podpora GPU výpočtov, ale aj portovanie na iOS, Android a FPGA.

Cielom frameworku Torch je mať maximálnu flexibilitu a rýchlosť pri vytváraní algoritmov a experimentov so zachovaním jednoduchosti vývojového procesu. Torch obsahuje rozsiahly ekosystém knižníc pre strojové učenie, počítačové videnie, spracovanie signálov, paralelné výpočty, spracovanie obrazu, audia a videa a i.

<sup>1</sup>The NVIDIA CUDA® Deep Neural Network library (cuDNN) <https://developer.nvidia.com/cudnn>

<sup>2</sup>BAIR <http://bair.berkeley.edu/>

<sup>3</sup>Yangqing Jia <http://daggerfs.com/>

<sup>4</sup>SuperVision model <http://www.image-net.org/challenges/LSVRC/2012/supervision.pdf>



Jadrom tohto frameworku sú ale knižnice pre neurónové siete a optimalizáciu. Možné je budovať komplexné topológie neurónových sietí a paralelizovať ich efektívne na CPU aj GPU [5].

## TensorFlow™

TensorFlow™ je knižnica s otvoreným zdrojovým kódom pre numerické počítanie s použitím diagramov dátových tokov. Uzly v diagrame reprezentujú matematické operácie, zatiaľ čo tok dát reprezentuje multidimenzionálne dátové polia tenzory. Flexibilná architektúra umožňuje rozdeliť výpočty na viacero CPU alebo GPU na PC, serveri, alebo mobilnom zariadení s jediným API. TensorFlow bol pôvodne vytvorený výzkumníkmi a inžiniermi Google Brain Team<sup>5</sup> výzkumnej organizácie Google's Machine Intelligence<sup>6</sup> pre účel riadenia strojového učenia a výzkumu hlbokých neurónových sietí, ale systém je dostatočne univerzálny aj pre použitie v rôznych iných oblastiach [4].

Medzi ďalšie frameworky pre tréning neurónových sietí patria aj Theano<sup>7</sup>, neon<sup>8</sup>, a ďalšie.

---

<sup>5</sup>Google Brain Team <https://research.google.com/teams/brain/>

<sup>6</sup>Google's Machine Intelligence <https://research.google.com/pubs/MachineIntelligence.html>

<sup>7</sup>Theano <http://deeplearning.net/software/theano/>

<sup>8</sup>neon <http://neon.nervanasys.com/docs/latest/>

## Kapitola 5

# Súčasný stav

Výpočtom hĺbkovej mapy sa dlhodobo zaoberá množstvo výskumných skupín. Vďaka tomu existujú rôzne prístupy a metódy na riešenie tejto problematiky.

Veľké množstvo existujúcich algoritmov prevádza postupnosť (podmnožinu) krokov:

1. výpočet párovacej ceny
2. agregácia ceny
3. výpočet disparity / optimalizácia
4. zdokonalenie disparity

Použité kroky sa ale môžu líšiť v závislosti od špecifického algoritmu.

Napr., *lokálne* algoritmy, v ktorých výpočet disparity v určitom mieste závisí len od intenzity pixlov daného okolia, väčšinou prevádzajú vyhladenie implicitne agregáciou ceny. Postup výpočtu niektorých z týchto algoritmov môže byť rozdelený na 3 kroky. Napr. SSD (sum-of-squared-differences) algoritmus môže byť popísaný:

1. párovacia cena je druhou mocninou rozdielu intenzít pri danej disparite
2. agregácia je prevádzaná aplikáciou pravouhlého okna na ceny s konštantnou disparitou
3. disparity sú vypočítané výberom minimálnej agregovanej hodnoty u každého pixelu

Niektoré ďalšie lokálne algoritmy však kombinujú krok 1 a 2 a používajú normalizovanú cross-koreláciu pre výpočet párovacej ceny.

*Globálne* algoritmy explicitne prevádzajú vyhladzovanie a optimalizáciu. Takéto algoritmy typicky vynechávajú agregačný krok, a riešia priradenie disparity (krok 3), ktoré minimalizuje globálnu cenovú funkciu, ktorá kombinuje dáta (krok 1) a podmienky vyhladzovania. Hlavným rozdielom týchto algoritmov je použitá minimalizačná procedúra.

Medzi týmito dvomi triedami algoritmov existujú aj iteratívne algoritmy, ktoré explicitne nešpecifikujú globálnu funkciu, ktorá má byť minimalizovaná, ale ktorých chovanie napodobňuje iteratívne optimalizačné algoritmy. Hierarchické algoritmy sa podobajú iteratívnym algoritmom, ale typicky pracujú s obrazovou pyramídou, kde sú výsledky z hrubších úrovní použité na vynútenie lokálnejšieho hľadania v jemnejších úrovniach [22].

Metóda	Priemerná chyba				
	$bad_{05}$	$bad_{10}$	$bad_{20}$	$bad_{40}$	$rms$
PMSC	39.1 <sub>1</sub>	14.8 <sub>2</sub>	6.71 <sub>3</sub>	4.44 <sub>4</sub>	12.9 <sub>1</sub>
LW-CNN	39.7 <sub>2</sub>	14.9 <sub>3</sub>	7.04 <sub>4</sub>	4.50 <sub>5</sub>	21.2 <sub>22</sub>
3DMST	39.9 <sub>3</sub>	14.5 <sub>1</sub>	5.92 <sub>1</sub>	3.72 <sub>1</sub>	13.3 <sub>2</sub>
MC-CNN+TDSR	40.1 <sub>4</sub>	16.1 <sub>5</sub>	6.35 <sub>2</sub>	3.81 <sub>2</sub>	16.8 <sub>11</sub>
MeshStereoExt	40.1 <sub>5</sub>	15.6 <sub>4</sub>	7.08 <sub>5</sub>	4.36 <sub>3</sub>	15.4 <sub>5</sub>
NTDE	40.4 <sub>6</sub>	16.2 <sub>6</sub>	7.44 <sub>7</sub>	4.55 <sub>6</sub>	15.7 <sub>8</sub>
MC-CNN-acrt	40.7 <sub>7</sub>	17.1 <sub>7</sub>	8.08 <sub>8</sub>	4.91 <sub>8</sub>	21.3 <sub>23</sub>
SNP-RSM	41.2 <sub>8</sub>	18.0 <sub>10</sub>	8.75 <sub>10</sub>	4.95 <sub>9</sub>	15.6 <sub>7</sub>
MCCNN_Layout	41.2 <sub>9</sub>	17.8 <sub>8</sub>	8.94 <sub>11</sub>	5.72 <sub>11</sub>	23.5 <sub>31</sub>
MC-CNN-fst	41.3 <sub>10</sub>	18.0 <sub>9</sub>	9.47 <sub>12</sub>	6.70 <sub>14</sub>	22.4 <sub>25</sub>

Tabuľka 5.1: Výpis niektorých z najúspešnejších metód v rebríčku *Middlebury Stereo Evaluation*. Hodnoty reprezentujú nameranú priemernú chybu pri zvolenej metrike. Číslo v indexe označuje poradie v rebríčku. Zdroj: <http://vision.middlebury.edu/stereo/eval3/>.

Výkonnostne najúspešnejšie existujúce metódy (tabuľka 5.1) sú porovnávané a vyhodnocované v online rebríčkoch, napr. *Middlebury Stereo Evaluation*<sup>1</sup> alebo *KITTI Stereo Evaluation*<sup>2</sup>.

Popisované algoritmy v tejto kapitole a aj navrhnutá metóda v tejto práci sú testované systémom na portáli Middlebury Stereo Evaluation. Metriky pre porovnanie vypočítaných disparitných máp sú popísané v kapitole 3.1.

## 5.1 Metóda PMSC

Jednou z najúspešnejších metód v rebríčku *Middlebury Stereo Evaluation* je metóda *PMSC - PatchMatch-Based Superpixel Cut for Accurate Stereo Matching*. V tejto metóde sa využívajú 3D značky, ktorými sa na miesto jednej hodnoty disparity pridelujú každému pixlu 3 hodnoty reprezentujúce disparitu a súčasne aj smer normály povrchu. Takýto prístup prináša niekoľko výhod. Zabráňujú vznikaníu aliasingu, ktorý sa často prejavuje na metódach, kde pixle obsahujú len hodnotu disparity. Korešpondujúce regióny na šikmých povrchoch sú jednoducho rozhodované pomocou povrchových normál.

Avšak kvôli obrovskému  $\mathbb{R}^3$  priestoru 3D značiek a nekonvexným funkciám energie, agregácia ceny ani globálna optimalizácia nie je taká zrejmá ako pri metódach so skalárnou disparitou. Pre vyriešenie týchto problémov je využitý algoritmus *PatchMatch based Superpixel Cut*.

V prístupe tejto metódy sa riešia dva problémy. Prvý spočíva vo výpočte párovacej ceny výrezov. Pre 1D disparity vytvorené zo štvorcových výrezov existuje množstvo postupov pre výpočet párovacej ceny, ale predpoklad konštantnej disparity v každom párovacom okne obmedzuje veľkosť okien na príliš malé pri slabej jedinečnosti. Použitím šikmých výrezov a 3D značiek sa prekonáva nevýhoda paralelných výrezov, takže veľkosti okien sú väčšinou výrazne väčšie. Avšak ako sa menia porovnávané okná pri rôznych 3D značkách, rozširovať existujúce postupy výpočtu párovacej ceny pre paralelné výrezy na šikmé výrezy je všeobecne neefektívne. Výsledkom toho sú výpočty párovacej ceny pre šikmé výrezy väčši-

<sup>1</sup>Middlebury Stereo Evaluation <http://vision.middlebury.edu/stereo/eval3/>

<sup>2</sup>KITTI Stereo Evaluation [http://www.cvlibs.net/datasets/kitti/eval\\_stereo.php](http://www.cvlibs.net/datasets/kitti/eval_stereo.php)

nou prevádzané agregáciou intenzity pixlov a zmenami gradientu, čo ale nie je použiteľné v mnohých situáciách ako napr. zmeny ožiarenia alebo osvetlenia medzi ľavým a pravým obrazom.

Pre vyriešenie tohto problému je použitý nový postup výpočtu ceny s dvojvrstvovou stratégiou. Najprv je predikovaná miera podobnosti malých paralelných okien natrénovanou konvolučnou neurónovou sieťou. Potom podobnosť pixlov je agregovaná adaptívnymi váhami na 3D značkových veľkých šikmých oknách. Prvá vrstva vylepšuje robustnosť lokálneho porovnávania na malých paralelných štvorcových oknách, kým druhá vrstva zabezpečuje konzistentnosť so šikmými povrchmi objektov na veľkých šikmých oknách.

Druhý problém riešený v tejto metóde je ako previesť globálnu optimalizáciu na nekonečnom trojdimenzionálnom značkovom priestore každého pixlu. Jedno efektívne riešenie je segmentácia. Po prevedení presegmentácie na obraze intenzít, metódy založené na segmentácii vypočítajú kandidátne 3D značky na superpixelovej úrovni s predpokladom, že pixle v rámci jedného segmentu patria k rovnakému 3D povrchu. Najlepšie priradenie je potom nájdené z konečnej množiny značiek miesto celého priestoru 3D značiek  $\mathbb{R}^3$ . Iná skupina metód využíva princíp PatchMatch pre aproximáciu najbližšieho suseda. PatchMatch Stereo redukuje obrovský prehľadávací priestor náhodným vzorkovaním a šírením suseda, a priamo priraduje aproximujúco najlepšiu 3D značku z  $\mathbb{R}^3$  pre každý pixel. Pridané sú aj regularizačné podmienky pre konštrukciu globálnych metód. Napriek dosiahnutiu vyššej presnosti, optimalizačné procesy globálnych PatchBased metód ľahko zlyhávajú pri falšných lokálnych minimách. Taktiež extrémne redundantné vzorkovanie a šírenie pre každý pixel spôsobuje neakceptovateľné výpočtové časy pri obrazoch s vysokým rozlíšením.

S využitím výhod efektívnosti metód založených na segmentácii a presnosti PatchMatch metód, je prezentovaná nová stratégia globálnej optimalizácie. Vzhľadom na pozorovanie, že vzorové hĺbkové mapy sú často priestorovo vyhladené s ohnivením na hranách objektov, je vykonávaná viacvrstvová superpixelová segmentácia a iteratívne šírená kandidátne značka superpixelmi. Náhodné zdokonalovanie PatchMatch metód je tiež použité pre aktualizáciu množín kandidátnych značiek každého pixlu, v ktorom je snaha nájsť lepšie priradenie celého obrazu [14].

Ukážka výsledku tejto metódy je na obr. 5.1.



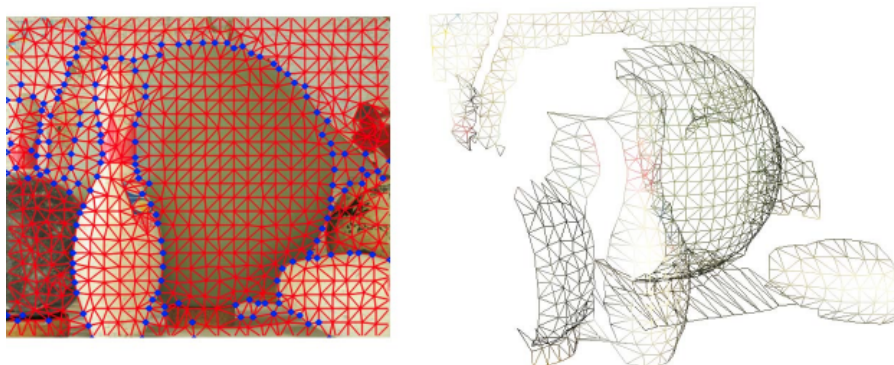
Obr. 5.1: Ukážka výstupu metódy PMSC. Obrazy zľava: ľavý vstupný obraz, pravý vstupný obraz, disparitná mapa (výstup metódy). Zdroj: <http://vision.middlebury.edu/stereo/eval3/>.

## 5.2 Metóda MeshStereoExt

Táto metóda využíva odlišný prístup, ako vyššie popísané algoritmy. Navrhovaný postup výpočtu hĺbkovej mapy v prvom kroku delí vstupný obraz na 2D trojuholníkové oblasti

podľa rozloženia hrán a lokálnej vizuálnej konzistencie, ako na obr. 5.2. Pre vygenerovanie korešpondujúceho 3D modelu je kľúčovým krokom správne spočítať disparity vrcholov jednotlivých trojuholníkov. Avšak kvôli nespojitosti hĺbky v obraze, vrcholy na hraniciach rôznych hĺbok by mali byť delené to iných trojuholníkov. Pre modelovanie deliacich vlastností je ku každému vrcholu priradená skrytá deliaca premenná. S dostupnými deliacimi vlastnosťami vrcholov môže byť regiónmi založený stereo prístup upravený pre optimalizáciu rovinnnej hĺbkovej mapy podľa toho, či prilahlé trojuholníky majú byť zarovnané, alebo rozdelené. Po vypočítaní hĺbkovej mapy môžu byť aktualizované deliace vlastnosti. Iteratívnym opakovaním týchto dvoch krokov sú získané optimalizované disparity vrcholov trojuholníkov.

Ukážka výsledku tejto metódy je na obr. 5.3.



Obr. 5.2: 2D triangulácia a 3D model generovaný metódou MeshStereo [27].



Obr. 5.3: Ukážka výstupu metódy MeshStereoExt. Obrazy zľava: ľavý vstupný obraz, pravý vstupný obraz, disparitná mapa (výstup metódy). Zdroj: <http://vision.middlebury.edu/stereo/eval3/>.

### 5.3 Metóda MC-CNN-acrt

Inšpirácia tejto práce, MC-CNN metóda, sa zameriava na prvý krok obecného postupu výpočtu disparitnej mapy, a to výpočet párovacej ceny. Problém je riešený tréňovaním konvolučných neurónových sietí pre porovnanie a odhad podobnosti dvoch malých obrazov. Tréňovanie je realizované vytvorením binárneho klasifikačného datasetu s entitami reprezentujúcimi podobné a rozdielne dvojice obrazov. Výstup konvolučnej neurónovej siete je použitý pre inicializáciu stereo párovacej ceny. Nasleduje postupnosť krokov: agregácia ceny, semiglobálne párovanie, kontrola konzistencie ľavej a pravej disparitnej mapy, mediánový a bilaterálny filter [28].

Ukážka výsledku tejto metódy je na obr. 5.4.



Obr. 5.4: Ukážka výstupu metódy MC-CNN-acrt. Obrazy zľava: ľavý vstupný obraz, pravý vstupný obraz, disparitná mapa (výstup metódy). Zdroj: <http://vision.middlebury.edu/stereo/eval3/>.



## Kapitola 6

# Návrh riešenia výpočtu mapy disparity zo stereo obrazu

Cieľom tejto práce je vytvorenie aplikácie, ktorá zo vstupného páru obrazov vypočíta disparitnú mapu a vytvorí jej vizuálnu reprezentáciu. Z predošlých kapitol vyplýva, že výpočet disparitnej mapy zo stereo obrazu nie je triviálny problém. V kapitole 5 sú popísané obecné princípy niektorých používaných metód. Na základe týchto princíпов a ďalších popísaných v [28] a [26], bolo riešenie zadaného problému navrhnuté nasledovným postupom:

1. Vytvorenie konvolučnej neurónovej siete pre výpočet miery podobnosti dvojice obrázkov.
2. Návrh stereo metódy – spracovanie výstupu neurónovej siete, výpočet surovej disparitnej mapy, agregácia a korekcia ceny a filtrácia výslednej disparitnej mapy.

### 6.1 Konvolučná neurónová sieť pre porovnanie dvojice obrázkov

Prvým krokom výpočtu disparitnej mapy je výpočet párovacej ceny blokov každého pixlu ľavého obrazu pre všetky možné úrovne disparity. Táto operácia sa prevádza pre získanie najpravdepodobnejšej hodnoty disparity v danej pozícii hľadaním najpodobnejšieho bloku v pravom obraze v smere epipolárnych čiar. Keďže však portál *Middlebury Stereo Evaluation* využíva pre testovanie rektifikované datasety, ktorých epipolárne čiary sú horizontálne zarovnané, postačuje hľadať podobné výrezy len v horizontálnom smere. Navyše, pre výpočet disparitnej mapy pre ľavý obraz stereo páru postačuje prehľadávať pozície v pravom obraze v intervale  $\langle x, x - d_{max} \rangle$ , kde  $x$  je horizontálna pozícia spracovávaného pixlu v ľavom obraze a  $d_{max}$  je maximálny počet úrovní disparít pre daný stereo pár.

Podľa [28] v návrhu som použil pre tento účel konvolučnú neurónovú sieť. Jej úlohou teda má byť porovnať výrezy z oboch obrazov a spočítať mieru podobnosti v %. Párovacia cena je potom definovaná ako

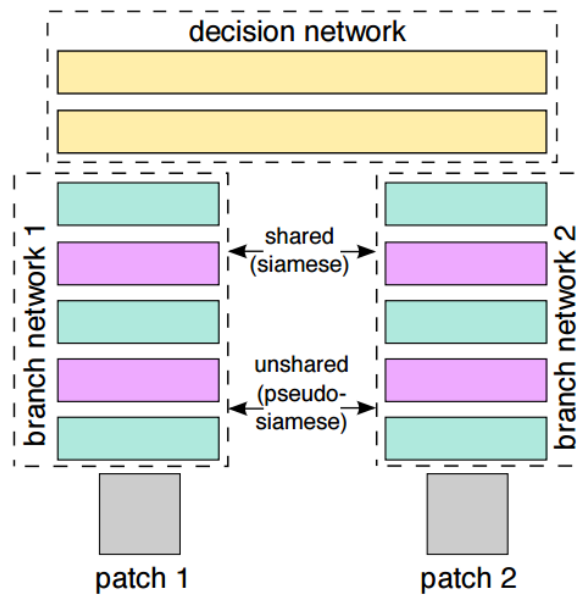
$$C_{CNN}(p, d) = -s(\langle \rho^L(p), \rho^R(p - d) \rangle), \quad (6.1)$$

kde  $-s(\langle \rho^L(p), \rho^R(p - d) \rangle)$  je výstup siete pre vstupné výrezy  $\rho^L(p)$  ľavého obrazu na pozícii  $p = [x, y]$  a  $\rho^R(p)$  pravého obrazu na pozícii  $p = [x - d, y]$ ,  $d$  označuje úroveň disparity.

Navrhol som neurónovú sieť siamskej architektúry 6.1. Sieť je zložená z dvoch vetiev, ktoré zdieľajú rovnaký súbor váh ako aj rovnakú architektúru. Každá vetva preberá na vstupe jeden z výrezov, a prevádza sériu konvolučných a ReLU vrstiev. Výstupy vetiev sú konkatenované, a posúvané plne prepojeným a ReLU vrstvám [26]. Výsledok je nakoniec normalizovaný funkciou sigmoidy. Poliacie vrstvy som vynechal z dôvodov popísaných v kapitole 7. Neurónovú sieť som navrhol ako binárny klasifikátor. Na vstup dostáva dvojicu šedotónových výrezov, ktorú percentuálne zaraďuje do dvoch tried: trieda rozdielnych obrazov a trieda zhodných obrazov. Súčet hodnôt vo výstupe siete pre vstupný pár je rovný 1, z toho dôvodu je pre ďalšiu prácu použitý len výstup pre triedu zhodných obrazov.

Vynechanie poliacich vrstiev implicitne zvyšuje počet potrebných konvolučných vrstiev, ktoré zas implicitne definujú rozmer vstupných obrazov. Veľký počet konvolučných vrstiev ale výrazne spomaľuje trénovací a výpočtový čas siete. Z toho dôvodu som navrhol siete pre vstupné výrezy s rozmermi od  $9 \times 9$  do  $15 \times 15$  pixlov. Siete sa líšia v počte konvolučných vrstiev. Návrh architektúry jednej z týchto sietí je na obr. 6.2.

Pre neurónovú sieť bolo potrebné navrhnuť taktiež trénovací dataset. Postupom inšpirovaným [28] som vytvoril binárny dataset, tvorený objektami, ktoré obsahujú dva obrazy. Tento objekt je zaradený do dvoch tried podľa toho, či má byť uložená dvojica posudzovaná ako podobná (trieda 1), alebo odlišná (trieda 0).



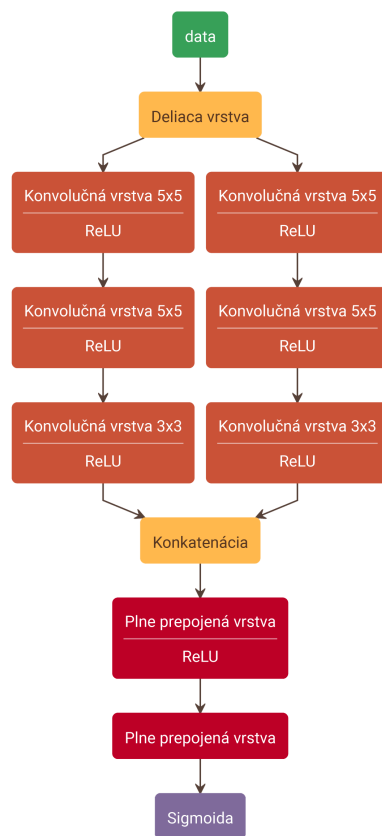
Obr. 6.1: Obecná architektúra siamskej konvolučnej neurónovej siete [26].

## 6.2 Stereo metóda

Pre demonštráciu výstupov a vplyvov jednotlivých častí metódy popisovanej v tejto práci som generoval niekoľko disparitných máp v priebehu spracovania jedného stereo páru.

Neurónová sieť produkuje pre každý pixel ľavého obrazu párovaciu cenu všetkých možných úrovní disparity pre daný stereo pár. Tieto dáta sú základom pre ďalší proces výpočtu mapy disparity.





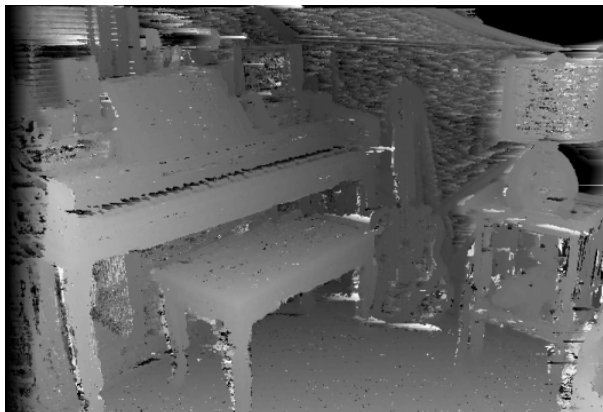
Obr. 6.2: Navrhnutá architektúra siamskej konvolučnej neurónovej siete pre vstupné obrázky s rozmermi  $11 \times 11$  pixlov.

Z využitím takýchto dát je možné vypočítať tzv. surovú disparitnú a to tak, že pre každý pixel ľavého obrazu je vybraná úroveň disparity s najnižšou cenou. Táto hodnota je normalizovaná maximálnym počtom úrovní disparity, a uložená ako intenzita vo výslednom obraze 6.3. Takáto mapa však obsahuje veľa nepresností. Navrhol som preto postup vylepšenia kvality disparitnej mapy inšpirovaný [28] v niekoľkých krokoch:

- Agregácia párovacej ceny pre vyváženie ceny v rámci pixlov prislúchajúcich rovnakému objektu.
- Korekcia najlepších úrovní disparity pre vylepšenie párovacej ceny.
- Agregácia párovacej ceny s cieľom odstránenia chybných rozhodnutí neurónovej siete a filtrácia výslednej disparitnej mapy.

### Agregácia ceny

Prvým krokom pre zlepšenie disparitnej mapy je agregácia párovacej ceny. Keďže agregácia ceny nad fixným okolím pixlov rapídne zlýháva v oblastiach skokových zmien disparít, bolo potrebné prevádzať agregáciu nad adaptívne zvoleným okolím každého pixlu. Preto pre každý pixel ľavého obrazu zostrojujem okolie zahrnutím pixlov s podobnou intenzitou s nádejou, že pixle patria rovnakému objektu.

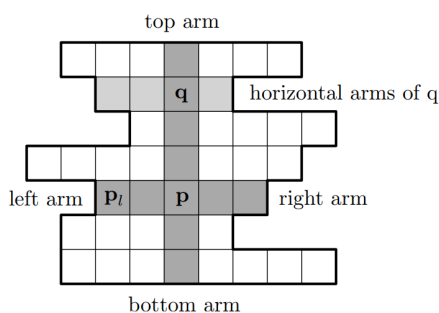


Obr. 6.3: Surová disparitná mapa spočítaná priamo z výstupov CNN.

Metóda začína budovaním ramien v horizontálnom a vertikálnom smere pre každý pixel obrazu. Rameno  $p_l$  v pozícii  $p$  sa v danom smere rozširuje, kým sú splnené podmienky:

- $|I(p) - I(p_l)| < agr_i$ ; rozdiel intenzít medzi pixlami ramena a ústredným pixlom by mal byť minimálny
- $\|p - p_l\| < agr_d$ ; vzdialenosť medzi bodmi ramena a ústredným bodom nesmie presiahnuť maximálnu veľkosť okolia

Skonstruovaním štyroch ramien vytváram kompletne okolie pixlov  $U(p)$  zjednotením horizontálnych ramien všetkých pozícií ležiacich na vertikálnych ramenách daného pixlu (6.4). Takéto okolia vytváram pre obidva obrázky vstupného stereo páru.



Obr. 6.4: Kompletne okolie bodu  $p$  je zjednotením horizontálnych ramien všetkých bodov  $q$  na vertikálnych ramenách bodu  $p$  [28].

Pre agregáciu ceny používam ale kombinované okolie  $U_d$ , ktoré vytváram prienikom kompletných okolí bodu  $(x, y)$  ľavého obrazu a bodu  $(x - d_{min}, y)$  pravého obrazu, kde  $d_{min}$  je úroveň disparity bodu  $(x, y)$  s najnižšou cenou, získanej z výstupov konvolučnej neurónovej siete:

$$U_d(p) = \{q | q \in U^L(p), q - d_{min} \in U^R(p - d_{min})\}. \quad (6.2)$$

Párovacia cena je potom priemerovaná nad kombinovaným okolím každého pixlu ľavého obrazu:

$$C_{AGR}^0(p, d) = C_{CNN}(p, d) \quad (6.3)$$

$$C_{AGR}^i(p, d) = \frac{1}{|U_d(p)|} \sum_{q \in U_d(p)} C_{AGR}^{i-1}(q, d) \quad (6.4)$$

kde  $i$  je číslo iterácie. Keďže sa okolia jednotlivých bodov prekrývajú, opakujem agregáciu niekoľko krát.

### Korekcia najlepších úrovní disparity

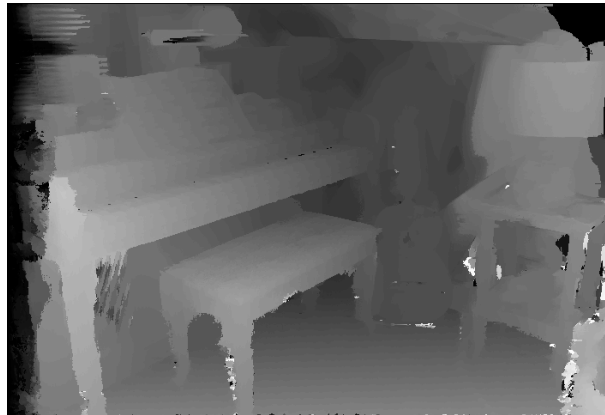
Ďalším krokom spracovania disparitynej mapy je korekcia najlepších úrovní disparity. Po agregácií párovacej ceny sa eliminoval šum a lokálne nepresnosti disparity.

Korekcia disparity  $C_{corr}$  prebieha vyhľadáním  $n$  úrovní disparít  $b_d$  s najlepšimi cenami. Tieto ceny sú korigované pomocou sumy absolútnych rozdielov  $SAD$  nad fixnými oknami v pozícii  $(x, y)$  v ľavom obraze a  $(x - d, y)$  v pravom obraze, pričom  $d$  je jedna z vybraných úrovní disparity:

$$C_{corr}(p, d) = \begin{cases} C_{AGR}^2(p, d)SAD(p, p - d), & \text{ak } d \in b_d(p, n) \\ 0, & \text{inak} \end{cases} \quad (6.5)$$

Ostatné úrovne disparít sú nastavené na 0.

Upravená mapa po prevedení agregácie ceny a jej korekcie je na obr. 6.5.



Obr. 6.5: Disparitná mapa po prevedení agregácie ceny dvomi iteráciami a korekciou prvých 12 najlepších úrovní disparity.

### Agregácia ceny a filtrácia

Záverecnými krokmi v spracovaní disparitynej mapy je opäť agregácia ceny jedinou iteráciou pre vylepšenie agregovaných oblastí nesprávne určených neurónovou sieťou, a následne vyhladenie vypočítanej disparitynej mapy bilaterálnym filtrom. Tento filter pomáha vylepšiť skokové prechody disparity na plynulejšie pri zachovaní ostrých výrazných hrán (obr. 6.6).



Obr. 6.6: Disparitná mapa po prevedení kompletnej stereo metódy.

## Kapitola 7

# Implementácia

Navrhnutý systém a nástroje pre výpočet disparitnej mapy som implementoval v jazykoch C++ a Python s využitím knižníc Caffé pre trénovanie neurónových sietí a OpenCV pre prácu s obrazom.

### Trénovacie datasety

Pre trénovanie sietí bolo potrebné vytvoriť datasety, ktoré sú tvorené anotovanými dvojicami obrázkov s požadovanými rozmermi. Skript `prepare_data.py` vyreže a uloží jednotlivé obrázky s veľkosťou  $64 \times 64$  pixlov z datasetov Liberty a Notre Dame. Skripty `create_train_liberty_hdf5.py` a `create_train_notredame_hdf5.py` vytvárajú trénovacie datasety a skript `create_validation_liberty_hdf5.py` testovací dataset anotovaných dvojíc s použitím vyrezaných obrázkov pomocou skriptu `prepare_data.py`. Vstupným parametrom jednotlivých skriptov je požadovaný rozmer obrázkov.

### Konvolučné neurónové siete

Navrhol a natrénoval som modely neurónových sietí so siamskou architektúrou bez poliacich vrstiev s použitím frameworku caffe. V adresároch `net_9`, `net_11` a `net_15` sú natrénované modely. Trénovanie prebiehalo s parametrami popísanými v tabuľke 7.1 na GPU.

### Program pre výpočet disparitnej mapy

S natrénovanými sieťami som implementoval stereo metódu. Hlavným súborom programu je `disparity_estimation.cu`. Načítava model siete a spúšťa sieť na celých šedotónových

type	SGD
base_lr	0.001
lr_policy	step
gamma	0.9
stepsize	10000
max_iter	400000
momentum	0.5
weight_decay	0.000001

Tabuľka 7.1: Parametre pri trénovaní neurónových sietí.

obrazoch vstupného stereo páru. Funkcie pre prácu s konvolučnou neurónovou sieťou sú implementované v súboroch `imagecompare.cpp` a `imagecompare.hpp`. Po získaní dát z neurónovej siete je v súbore `disparity_estimation.cu` spúšťaná stereo metóda. Metódu som implementoval v súboroch `postprocessing.h` a `postprocessing.cu` a niekoľkých ďalších súboroch. Spočítava okolia pixlov pre agregáciu ceny s použitím platformy CUDA v súboroch `combined_support_region.cu`, `construct_arms.cu` a `support_region.cu`, a prevádza kroky popísané v návrhu stereo metódy.

## Optimalizácia výpočtu disparitnej mapy

Pri štandardnom postupe by sa zo vstupných obrazov vyrezávali malé výrezy, ktoré sa sieťou vyhodnotia, uloží sa výsledok a proces sa zopakuje pre všetky pozície a všetky možné úrovne disparity vstupného páru obrazov. Takýto postup je ale časovo nepriateľne náročný, kvôli čomu som pri návrhu modelov sietí vynechal poliace vrstvy. Do takýchto modelov je možné na vstup poslať pár celých obrazov, ktoré sú konvolučnými vrstvami spracované, a na výstupoch posledných konvolučných vrstiev každej vetvy siamskej siete je  $hw$  vektorov dĺžky  $n$ , kde  $n$  je počet filtrov poslednej konvolučnej vrstvy a  $h$  a  $w$  sú rozmery vstupného obrazu. Tu sa prejavuje výhoda siamskej architektúry, kde je možné z posledných vrstiev jednotlivých vetiev vybrať vektory popisujúce jednotlivé výrezy v obraze, uložiť ich po stĺpcoch (column-major order), a znovu použiť pri ďalších výpočtoch. Keďže je potrebné prevádzať výpočet sieťou pre všetky úrovne disparity, pri uložených vektoroch posledných konvolučných vrstiev, ktoré sú spočítané len raz pre jeden stereo pár, je takýto prístup rapídne rýchlejší. Následne pre každú úroveň disparity  $d$  zmeňujem vstupné obrazy o jeden stĺpec zľava pre ľavý vstupný obraz, resp. sprava pre pravý vstupný obraz. Výpočet siete ale spúšťam až od vrstvy konkatenácie, na ktorej vstup privádzam uložené vektory s offsetom  $dh$  pre vektory ľavého obrazu, a  $h(w - d)$  vektorov pre pravý obraz.

Výsledkom takéhoto postupu je, že pre celý stereo pár sú konvolučné vrstvy neurónovej siete, ktoré sú najpomalšou časťou, spúšťané len raz, a plne prepojené vrstvy toľko krát, koľko je maximálnych možných úrovní disparity pre daný pár.

## Kapitola 8

# Experimenty a výsledky

Vzhľadom na to, že vytvorená aplikácia sa skladá z dvoch hlavných častí, konvolučnej neurónovej siete pre porovnanie výrezov obrazov a stereo metódy, bolo potrebné previesť experimenty a testy na aplikácii ako celku, ale aj na neurónovej sieti samostatne.

### 8.1 Konvolučná neurónová sieť pre porovnanie dvojice obrázkov

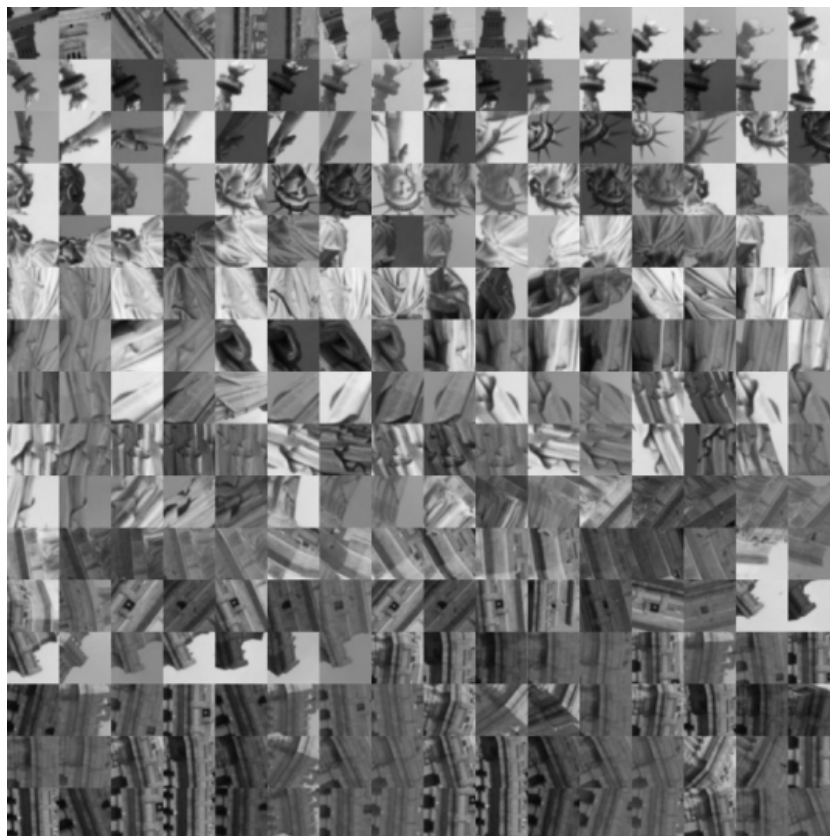
Cieľom návrhu, implementácie a testovania konvolučných neurónových sietí bolo získať model s čo najvyššou presnosťou pri porovnávaní dvoch výrezov obrazu, ktorý bude použitý vo výslednej aplikácii.

#### Použitý dataset

Dôležitou časťou návrhu a realizácie siete bol výber tréningového a testovacieho datasetu. Po analýze možných riešení tohto problému, teda vytvorenie vlastného datasetu zo stereo párov a ich vzorových disparitných máp alebo použitie iného existujúceho datasetu, som zvolil dataset, ktorý bol taktiež použitý pre tréningovanie siete v [26], a teda dataset z [7], ktorý pozostáva z troch sád, Yosemite, Notre Dame, a Liberty, z čoho každý obsahuje viac ako 450000 výrezov ( $64 \times 64$  pixlov). Každá sada bola generovaná s použitím 3D korešpondencií získaných z hĺbkových máp. Tieto mapy boli použité pre vytvorenie 500000 anotovaných párov výrezov pre každý dataset, s rovnakým počtom zhodných a odlišných prípadov. Pre tréningovanie siete som použil sady Liberty a Notre Dame. Na obr. 8.1 je ukážka tréningových dát a obr. 8.2 zobrazuje anotované páry použité pri tréningu siete.

#### Tréningovanie siete

Natrénoval som celkovo 3 siete pre porovnanie obrázkov malých rozmerov. Siete sa líšia v počte konvolučných vrstiev, ktoré implicitne udávajú veľkosť vstupných obrazov. Siam-skú architektúru som zachoval pri všetkých sieťach z optimalizačných dôvodov popísaných v kapitole 7. Tréningový progres jednotlivých sietí je zobrazený v grafe 8.3. Pre jednoduchšie odkazovanie na jednotlivé siete zavádzam nasledujúce označenia: `net_9` pre CNN so vstupnými obrazmi s rozmermi  $9 \times 9$  px., `net_11` pre CNN so vstupnými obrazmi s rozmermi  $11 \times 11$  px., `net_15` pre CNN so vstupnými obrazmi s rozmermi  $15 \times 15$  px. Ostatné parametre, ako počet konvolučných vrstiev či ich rozmery sú popísané v tabuľke 8.1.



Obr. 8.1: Podmnožina tréovacích dát Liberty pre konvolučnú neurónovú sieť pre porovnanie obrázkov. Jednotlivé políčka sú párované pomocou kalibračného súboru. Zdroj: <http://www.cs.ubc.ca/~mbrown/patchdata/patchdata.html>.






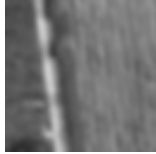


Sieť je trévaná ako binárny klasifikátor, a výstupom siete pre jednu dvojicu obrázkov, je percentuálne skóre pre dve triedy (trieda rozdielnych a trieda zhodných obrazov). Keďže súčet týchto skóre je rovný 1, v aplikácii je zohľadňované len skóre pre triedu zhodných obrazov, ktoré teda reprezentuje mieru podobnosti dvoch obrazov. Tabuľka 8.2 zobrazuje rozhodnutie jednotlivých sietí o podobnosti dvoch obrázkov. Siete dokázali rozlišovať rozdielne obrázky, ale percento podobnosti pridelené jednotlivým dvojiciam je diskutabilné. Podľa obr. 8.3 bol predpoklad, že najúspešnejšou sieťou bude `net_15`, ktorá dosahovala úspešnosti na testovacom datasete viac ako 93%, avšak ostatné siete výrazne nezaostávali. Ukázalo sa však, že kontrolovať subjektívne percentuálne rozhodnutie siete subjektívnym názorom človeka nie je vhodným spôsobom, ako preveriť správnu funkčnosť natrénovanej siete. Z toho dôvodu som funkčnosť sietí overil vygenerovaním nespracovanej disparitnej mapy. Proces získania takejto mapy je popísaný v kapitole 6. Nespracované výstupy sietí sú na obr. 8.4. Napriek tomu, že mapy nie sú dokonalé, obsahujú množstvo nepresností a šumu, je zrejme, že neurónové siete boli trévané pre tento účel správnym spôsobom a výrezy obrazov porovnávajú do určitej miery správne.

Pri surových disparitných mapách je potrebné si uvedomiť vplyv rozmeru vstupných dát na výsledný obraz. Najvýraznejšie rozdiely je možné pozorovať na súvislých plochách a pri hranách objektov, teda pri teoretickom vysokom rozdielne susedných disparít. Sieť s menšími vstupnými obrázkami generuje výrazne viac šumu, no na druhej strane lepšie



Parameter	net_9	net_11	net_15
veľkosť vstupných obrazov	$9 \times 9$	$11 \times 11$	$15 \times 15$
počet konvolučných vrstiev	2	3	4
veľkosť konvolučnej vrstvy #1	$5 \times 5$	$5 \times 5$	$5 \times 5$
veľkosť konvolučnej vrstvy #2	$5 \times 5$	$5 \times 5$	$5 \times 5$
veľkosť konvolučnej vrstvy #3	–	$3 \times 3$	$5 \times 5$
veľkosť konvolučnej vrstvy #4	–	–	$3 \times 3$

Tabuľka 8.1: Parametre tréovaných konvolučných neurónových sietí.

Vstup 1	Vstup 2	Miera podobnosti [%]		
		net_9	net_11	net_15
		78.9	96.9	98.6
		55.4	92.3	27.4
		7.4	0.4	0.0
		98.0	99.3	99.8

Tabuľka 8.2: Rozhodnutie tréovaných sietí o miere podobnosti pre dvojice obrázkov.



Obr. 8.2: Ukážka anotovaných dvojíc, ktoré som použil pri tréovaní a testovaní neurónovej siete. Vizálne podobné obrázky som radil do triedy 1, rozdielne do triedy 0.

zvláda zachovávať detaily, hlavne už spomenuté hrany objektov. Obr. 8.5 zobrazuje detailne tieto javy.

S chybami pri výpočte takejto dispartinej mapy sa počítalo, preto je vo výslednej aplikácii navrhnuté a implementované spracovanie výstupných dát z konvolučnej neurónovej. Ukázalo sa ale, že neurónové siete sú vhodným základom pre výpočet optimálnej dispartinej mapy. Je ale potrebné navrhnuť najvhodnejšiu architektúru a tréovať na veľmi veľkom datase, aby boli dosiahnuté najlepšie výsledky.

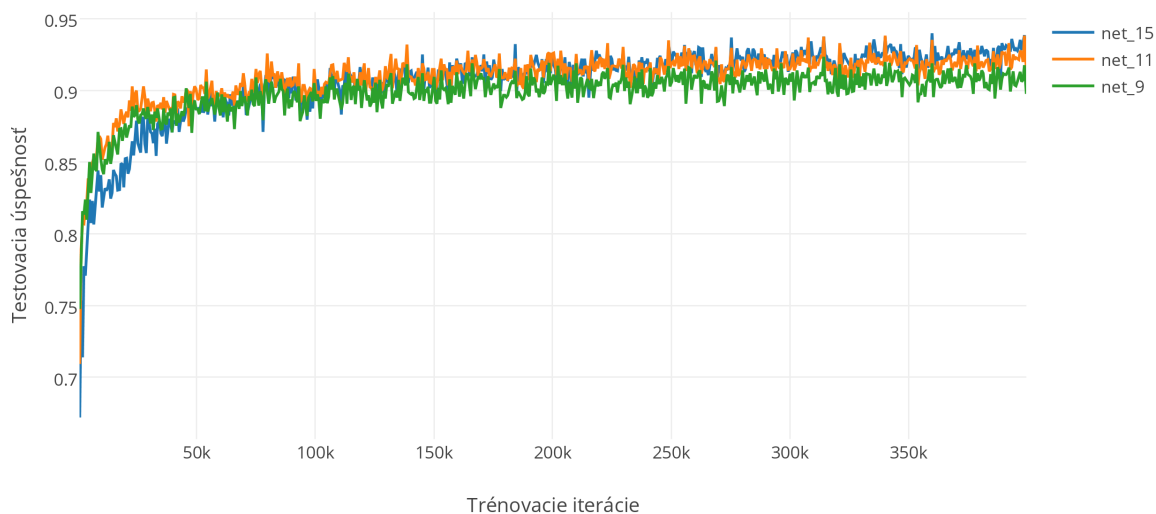
## 8.2 Stereo metóda

Z natréovaných sietí je ťažké vybrať najlepšiu, keďže každá má výhody aj nevýhody. Z toho dôvodu som testoval stereo metódu s využitím všetkých sietí. Dispartitnú mapu spočítanú len pomocou neurónovej siete som vylepšoval niekoľkými spôsobmi.

Keďže surová dispartitná mapa obsahuje veľké množstvo šumu ale aj výrazné do určitej miery korektné hrany objektov, ako prvú a najjednoduchšiu techniku spracovania výstupu z neurónovej siete som použil bilaterálny filter. Na obr. 8.6 je vidieť výsledok aplikácie tohto filtra. Filter dokázal vyhladiť obrázky, potlačiť jemný šum, no väčšie regióny, vzniknuté nesprávnym určením hodnoty disparity väčšieho množstva susediacich pixlov ostali nezlepšené. Taktiež pre zlepšenie hrán objektov nemal filter žiadny efekt.

Bilaterálny filter, ako sa ukázalo, vizuálne kvalitu dispartitných máp výrazne nezlepšil. Preto bolo potrebné navrhnuť a implementovať metódu, ktorá by dosahovala výrazne lepšie výsledky. Navrhol som stereo metódu, založenú na dvojfázovej agregácii párovacej ceny, ktorá je popísaná v kapitole 6. Keďže postup spracovania dispartitnej mapy tejto metódy prebieha tak, že je najprv aplikovaná agregácia párovacej ceny, potom je prevádzaná korekcia možných úrovní disparity s najlepšimi cenami pre každý pixel, a následne je opäť aplikovaná agregácia ceny nasledovaná bilaterálnym filtrom, demonštrujem vplyv metódy na výsledný obraz v dvoch okamihom procese. Konkrétne, po korekcii možných úrovní disparity pre každý pixel, a po bilaterálnom filtri, teda po skončení spracovacieho procesu. Obr. 8.7 a obr 8.8 zobrazuje dispartitné mapy v dvoch okamihoch spracovania, resp. detaily dispartitných máp po týchto fázach.

Vplyv spracovania popisovanej metódy na pôvodnú dispartitnú mapu z globálneho pohľadu je vidieť vo výraznom potlačení šumu a vyhladení obrazu. Spracované mapy neobsahujú také množstvo skokov v hodnotách disparity pri susedných pixloch ako pôvodné mapy. Pri detailnom pohľade sú ale vidieť konkrétnejšie vplyvy metódy. Vo výstupoch po fáze agre-



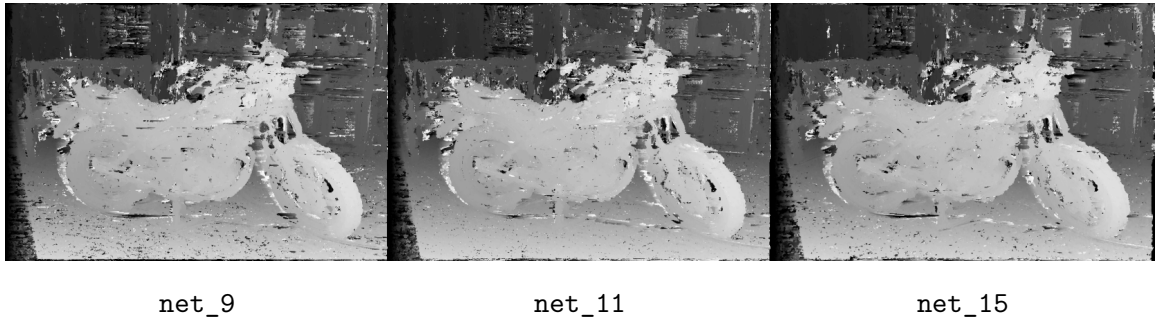
Obr. 8.3: Proces učenia konvolučných neurónových sietí. Je možné vidieť, že úspešnosť sa na testovacích dátach zvyšuje s časom a siete sa teda na použítom datasete učia správne rozhodovanie problému.

gácie a korekcie ceny je vidieť, že hrany objektov sú výrazne deformované, hlavne v oblasti kolesa motorky, ale aj pri výfukovom potrubí, aj keď v menšej miere. Na druhej strane, skupiny pixlov s nesprávne určenými hodnotami disparity sa vyskytujú vo výrazne menšej miere, a teda potlačenie takéhoto šumu a výlepšenie obrazu je zreteľné. Následné agregovanie ceny a bilaterálny filter, ako je vidieť, korigujú hodnoty niektorých pixlov tak, aby splývali s okolím. Zaujímavé je, že po tejto fáze sa v niektorých oblastiach obrazu zlepšili hrany objektov. Tento efekt je vidieť na výfukovom potrubí, ale aj v malej oblasti pri pravej časti kolesa. Takéto zlepšenie je najvýraznejšie pozorovateľné u **net\_15**. Bilaterálny filter na koniec vyhladzuje výslednú mapu a zjemňuje prechody hodnôt disparity.

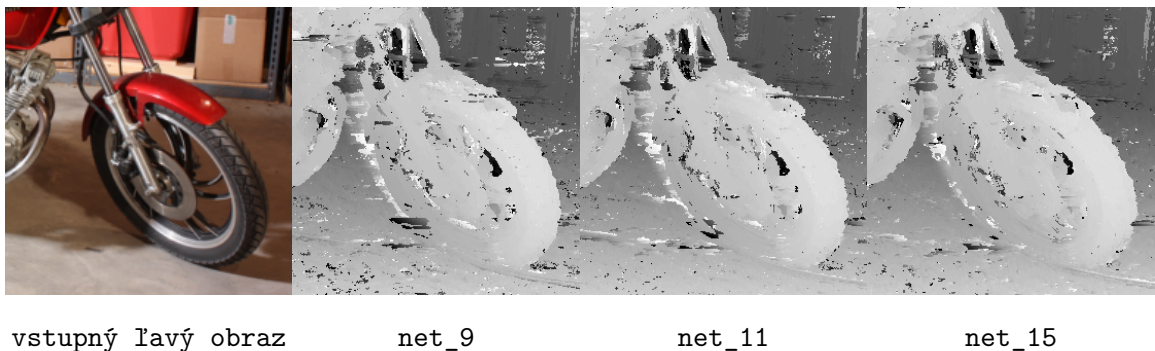
Keďže spracované disparitné mapy zo všetkých sietí sa vizuálne navzájom príliš nelíšia, pre ukážku výstupných disparitných máp rôznych vstupných obrazov a ich vzorových máp na obr. 8.9 som vybral výstupy pre metódu, ktorá využíva **net\_11**. Mapy na tomto obrázku sú prevedené do iného farebného priestoru, kde najteplejšie farby reprezentujú najvyššie hodnoty disparity.

### Porovnanie disparitných máp

Pri porovnávaní disparitných máp bolo dôležité zvoliť vhodné spôsoby a metriky pre posúdenie kvality výstupných obrazov. Portál *Middlebury Stereo Evaluation* poskytuje skripty pre generovanie potrebných súborov a dát pre porovnanie systému generujúcho disparitné mapy s ostatnými metódami v ich aktuálnom rebríčku. Taktiež dáva k dispozícii trénovacie datasety v rôznych rozlíšeniach: plné rozlíšenie F, polovičné rozlíšenie H a štvrtinové Q. Z dôvodu vysokých pamäťových nárokov implementovaného riešenia som disparitné mapy vypočítaval s použitím štvrtinového rozlíšenia datasetu. Metriky pre testovanie kvality dis-



Obr. 8.4: Nespracované disparitné mapy vypočítané pomocou konvolučných neurónových sietí. Najvyššia intenzita označuje najvyššiu hodnotu disparity, a teda najmenšiu vzdialenosť od kamery.



Obr. 8.5: Vplyv veľkosti vstupných výrezov na výsledok výpočtu CNN. Vidieť detailnejšie ostré hrany pri výfukovom potrubí motorčky či okraji kolesa u `net_9`. Naopak šum v pozadí je najslabší u `net_15`.

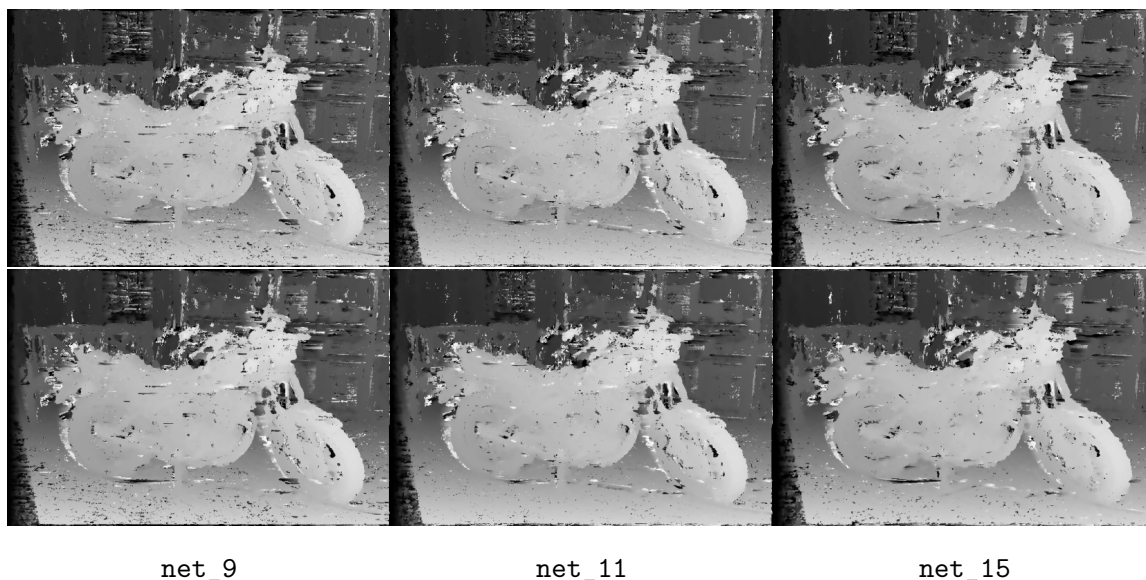
paritných máp som zvolil RMS chybu ( $rms$ ) a percento chybné spárovaných pixlov ( $bad_x$ ) s toleranciou chyby v štyroch úrovniach. Jednotlivé metriky popisujú vzorce 3.2 a 3.3.

Systém som testoval s využitím superpočítača virtuálnej organizácie MetaCentrum<sup>1</sup> na clustri doom<sup>2</sup>. Hardware na clustri doom zahŕňa procesor *2x 8-core Intel Xeon E5-2650v2 2.60GHz* a grafické karty *2x nVidia Tesla K20 5GB*.

Disparitné mapy po spracovaní sú síce vizuálne krajšie, avšak metriky pre ohodnotenie kvality máp zohľadňujú iné aspekty. Pre demonštráciu zlepšenia kvality popisovanou metódou som pomocou *Middlebury Stereo Evaluation* vyhodnotil niekoľko disparitných máp pre každú natrénovanú sieť. Popis výstupov je v tabuľke 8.3. Namerané hodnoty pre jednotlivé siete a výstupy sú v tabuľke 8.4. Číslo v indexe pri hodnotách udáva umiestnenie v rebríčku pri maximálnom počte 62 metód. Zvýraznené sú najlepšie výsledky v rámci danej metriky. Z uvedených hodnôt vyplýva, že po filtrácii surovej disparitnej mapy vypočítanej z výstupov konvolučnej neurónovej siete bilaterálnym filtrom vznikajú kvalitnejšie disparitné mapy podľa metrick  $bad_{05}$  a  $bad_{10}$ , ako s použitím kompletnej metódy. Spôsobené to môže byť aj tým, že pri agregácii ceny okrem deformácie hrán zanikajú aj plynulé prechody medzi úrovňami disparity toho istého objektu. Potvrdzujú to metriky vyššej tolerancie  $bad_{20}$  a  $bad_{40}$ , kde kompletná metóda dosahuje výrazne lepšie skóre, keďže sú viac

<sup>1</sup>MetaCentrum <https://metavo.metacentrum.cz/cs/>

<sup>2</sup>Cluster doom <https://metavo.metacentrum.cz/pbsmon2/resource/doom.metacentrum.cz>



Obr. 8.6: Jednoduché spracovanie surových disparitných máp aplikáciou bilaterálneho filtra. Hore pôvodné disparitné mapy, dole po aplikácii filtra.

d_raw	surová disparitná mapa
d_bilateral	čiastočne spracovaná mapa bilaterálnym filtrom
d_agregated	mapa vylepšená agregáciou ceny a jej korekciou
d_final	mapa spracovaná kompletnou metódou popísanou v tejto práci

Tabuľka 8.3: Popis jednotlivých výstupov v priebehu spracovania disparitnej mapy.

tolerované súvisé plochy jednej hodnoty disparity namiesto plynulého prechodu, naopak mape spracovanej len bilaterálnym filtrom škodí výrazný šum. Tento fakt dokazuje aj obr. 8.10. Napriek tomu je ale vidieť, že metóda vylepšuje pôvodné disparitné mapy, keďže sa chyba znižuje u všetkých metrik.

Z tabuľky je taktiež vidieť podľa umiestnenia, že navrhnutá a implementovaná metóda generovala kvalitnejšie mapy ako niektoré publikované metódy v spomínanom rebríčku. Konkrétne porovnanie s metódami *MC-CNN-acrt* [28], *ICSG* [24] a *SED* [8] je na obr. 8.11 a v tabuľke 8.5 s použitou metrikou  $bad_{40}$ . Pre porovnanie som vybral variantu so sieťou *net\_9*, ktorá dosahovala najlepšie výsledky spomedzi troch trébovaných. Výsledky nie sú také dobré, ako výsledky riešenia *MC-CNN-acrt*, z ktorého vychádza navrhnutá metóda v tejto práci. Môže to byť spôsobené rôznymi faktormi, od iného trébovacieho datasetu neurónovej siete, cez rozmer vstupného stereo páru až po samotnú stereo metódu, ktorá v tejto práci neobsahuje krok minimalizácie funkcie energie.

Testovaním sa ukázalo, že implementovaná metóda generuje disparitné mapy do určitej miery odpovedajúce skutočným. Najlepšie výsledky som dosiahol s použitím siete *net\_9*, ktorá zachováva najviac detailov. Avšak v kombinácii s inou sieťou s väčším rozmerom vstupných dát, ktoré dosahujú lepšie výsledky pri súvislých, nízko texturovaných oblastiach, by mohla byť kvalita disparitných máp výrazne väčšia.

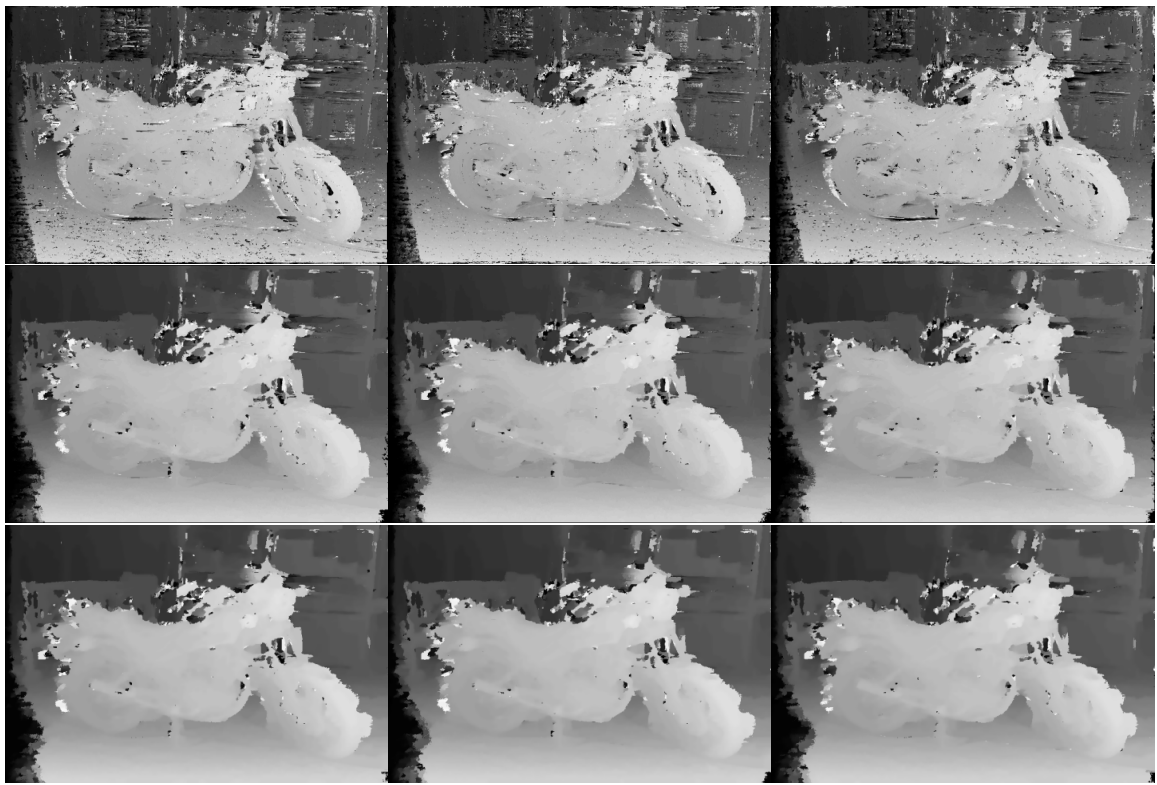
Použitá sieť	Výstup	$bad_x$ [%]				$rms$ [px]
		0.5	1.0	2.0	4.0	
net_9	d_raw	87.7 <sub>61</sub>	73.9 <sub>61</sub>	56.3 <sub>61</sub>	41.4 <sub>61</sub>	50.5 <sub>62</sub>
	d_bilateral	81.4 <sub>56</sub>	64.7 <sub>61</sub>	50.3 <sub>61</sub>	38.7 <sub>61</sub>	48.9 <sub>61</sub>
	d_agregated	85.6 <sub>61</sub>	69.3 <sub>61</sub>	48.0 <sub>61</sub>	30.9 <sub>59</sub>	36.8 <sub>59</sub>
	d_final	83.6 <sub>61</sub>	65.1 <sub>61</sub>	44.4 <sub>61</sub>	29.0 <sub>59</sub>	31.1 <sub>57</sub>
net_11	d_raw	87.5 <sub>61</sub>	73.4 <sub>61</sub>	55.0 <sub>61</sub>	39.4 <sub>61</sub>	48.1 <sub>61</sub>
	d_bilateral	80.2 <sub>55</sub>	62.6 <sub>59</sub>	47.8 <sub>61</sub>	36.4 <sub>61</sub>	46.5 <sub>61</sub>
	d_agregated	86.0 <sub>61</sub>	70.0 <sub>61</sub>	49.1 <sub>61</sub>	31.6 <sub>59</sub>	37.4 <sub>59</sub>
	d_final	84.1 <sub>61</sub>	66.0 <sub>61</sub>	45.4 <sub>61</sub>	29.7 <sub>59</sub>	32.4 <sub>58</sub>
net_15	d_raw	90.4 <sub>62</sub>	79.1 <sub>62</sub>	64.5 <sub>61</sub>	50.1 <sub>61</sub>	49.4 <sub>62</sub>
	d_bilateral	84.3 <sub>61</sub>	70.1 <sub>61</sub>	57.9 <sub>61</sub>	47.4 <sub>61</sub>	48.1 <sub>61</sub>
	d_agregated	86.6 <sub>61</sub>	71.4 <sub>61</sub>	51.1 <sub>61</sub>	33.3 <sub>60</sub>	39.0 <sub>59</sub>
	d_final	84.3 <sub>61</sub>	66.6 <sub>61</sub>	46.4 <sub>61</sub>	30.9 <sub>59</sub>	33.5 <sub>58</sub>

Tabuľka 8.4: Porovnanie sietí a rôznych výstupov metódy spracovania disparitnej mapy. Hodnoty v jednotlivých metrikách sú spočítané ako priemer výsledkov z celého testovacieho datasetu.

Metóda	$avgbad_{40}$	Dataset				Poradie
		Jadeplant	Motorcycle	Piano	Playroom	
MC-CNN-acrt	6.34	11.1	2.14	7.97	7.61	11.
moja	29.0	37.7	19.3	26.1	31.3	59.
ICSG	32.0	38.9	22.4	32.4	41.4	60.
SED	56.2	71.1	48.3	66.1	56.2	62.

Tabuľka 8.5: Porovnanie kvality disparitných máp rôznych metód. Priemerná hodnota zo všetkých testovacích datasetov  $avgbad_{40}$  je spočítaná z kompletného datasetu dostupného na stránke rebríčka.



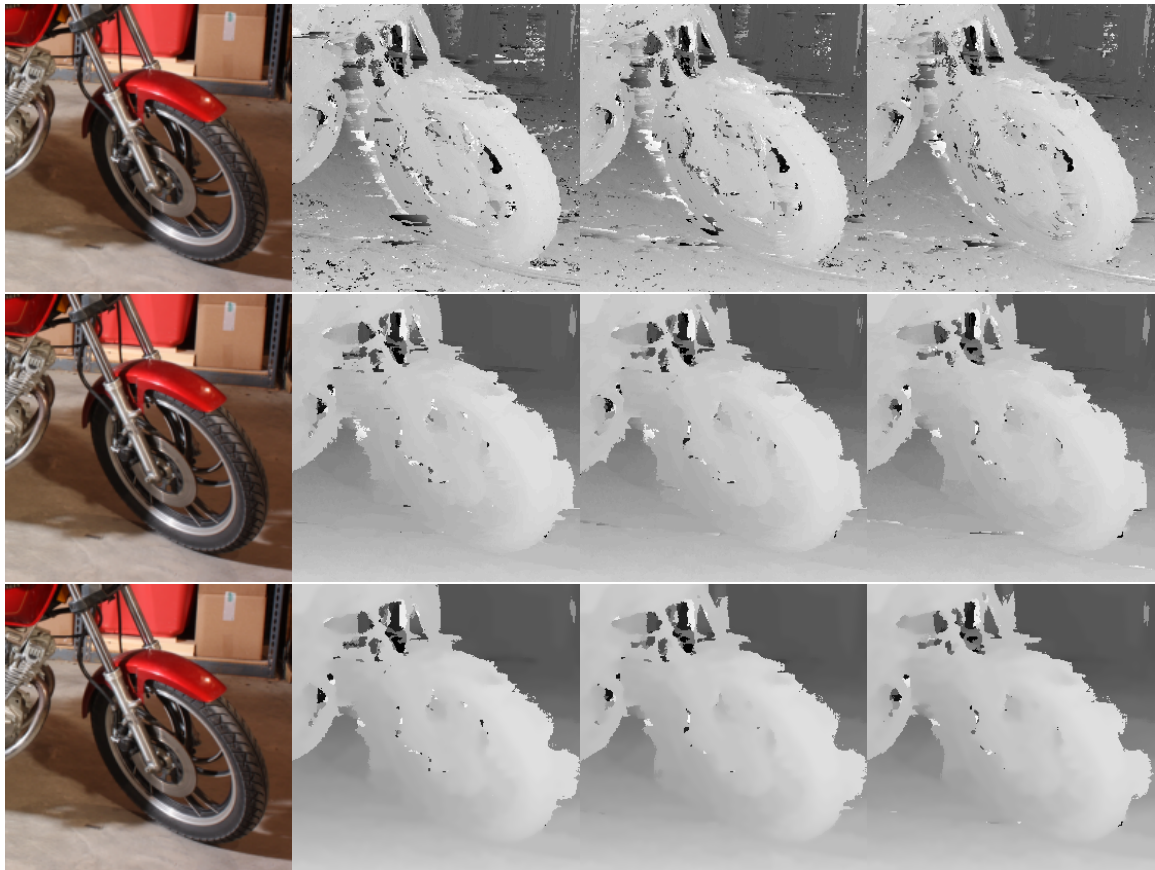


net\_9

net\_11

net\_15

Obr. 8.7: Disparitné mapy spracované agregáciou a korekciou ceny (v strede) a kompletnou metódou popisovanou v tejto práci (dole).



vstupný ľavý obraz

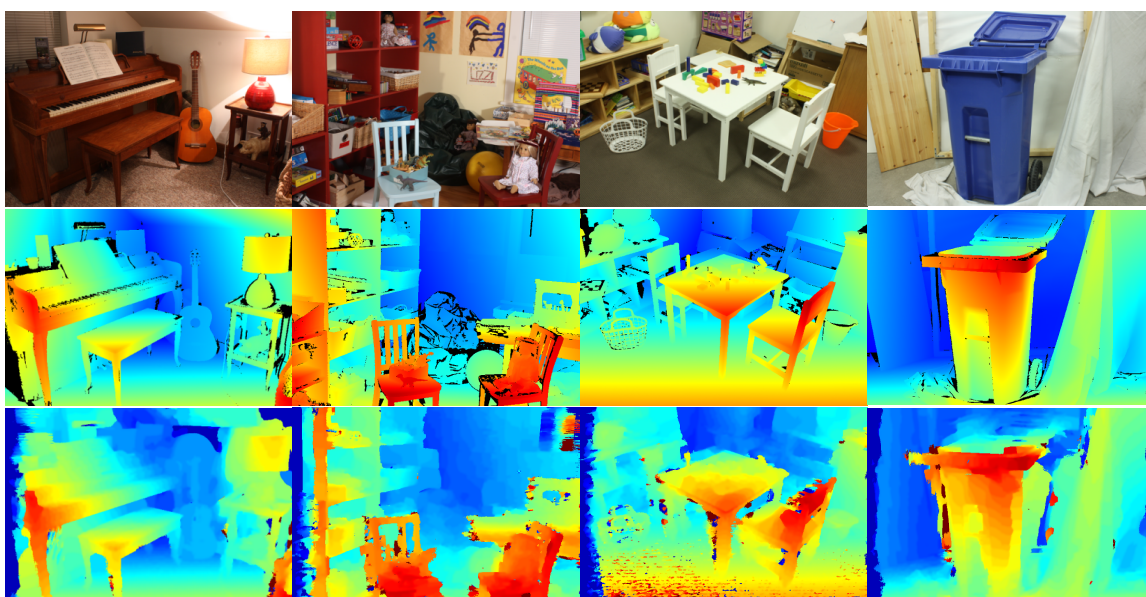
net\_9

net\_11

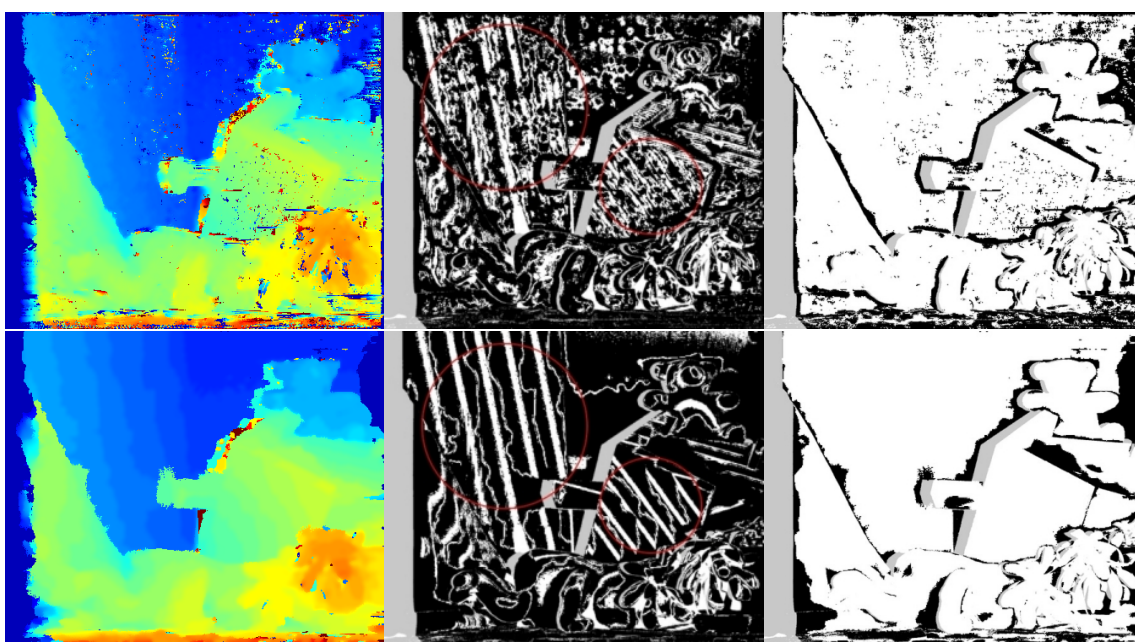
net\_15

Obr. 8.8: Detail disparitných máp spracovaných agregáciou a korekciou ceny (v strede) a kompletnou metódou popisovanou v tejto práci (dole).

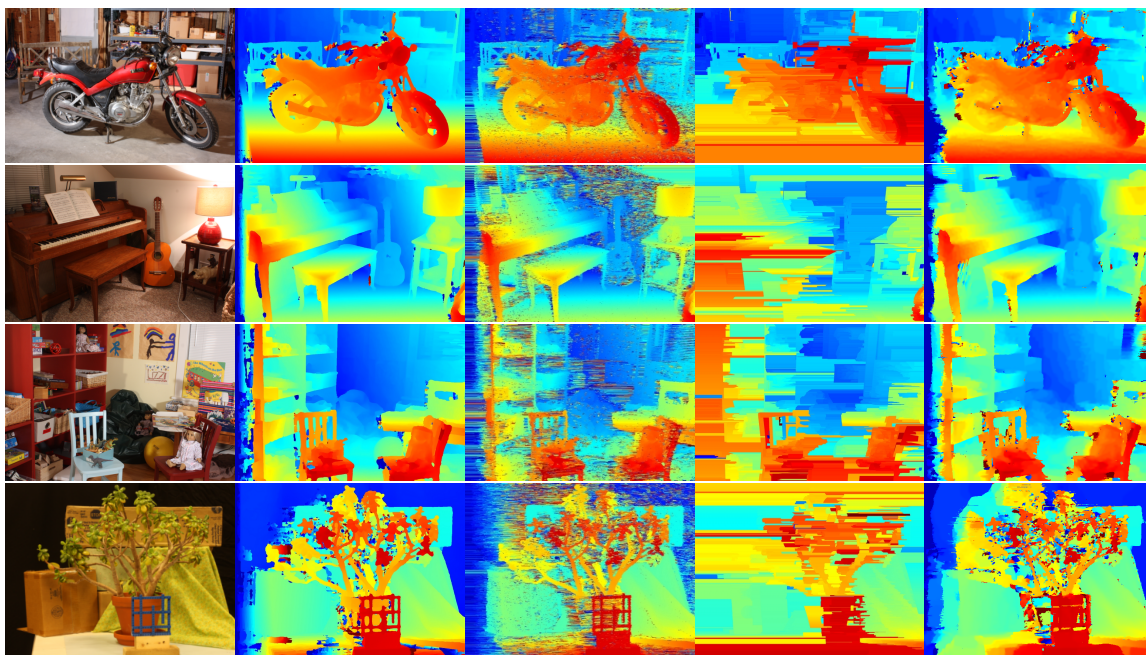




Obr. 8.9: Vstupné lavé obrazy (hore), vzorové disparitné mapy (v strede) a výstupné spracované disparitné mapy metódy popisovanej v tejto práci s využitím `net_11` (dole).



Obr. 8.10: Zľava: výstupná disparitná mapa (hore aplikovaný bilaterálny filter, dole kompletná metóda); mapa chybných pixlov, ktorých chyba je väčšia ako 0.5 ( $bad_{05}$ ); mapa chybných pixlov, ktorých chyba je väčšia ako 4.0 ( $bad_{40}$ ). U metriky  $bad_{05}$  je vo vyznačených oblastiach vidieť väčšie množstvo správne určených pixlov (biela) u bilaterálneho filtra, zatiaľ čo u kompletnej metódy sú vidieť výrazné pásy s nesprávne určenými pixlami medzi mini (čierna). To je spôsobené agregáciou ceny a s tou spojeným zaniknutím plynulých prechodov disparity. U metriky  $bad_{40}$  sa naopak skokové prechody tolerujú, a je vidieť, že kompletná metóda dosahuje výrazne nižšej chyby.



Obr. 8.11: Ukážky disparitných máp. Datasetsy zhora: Motorcycle, Piano, Playroom, Jadeplant. Metódy zľava: *MC-CNN-acrt*, *ICSG*, *SED* a moja. Zdroj: <http://vision.middlebury.edu/stereo/submit3/upload.php>.

## Kapitola 9

# Záver

Cieľom tejto práce bolo vytvoriť systém pre výpočet disparitnej mapy zo stereo obrazu s využitím konvolučných neurónových sietí pre výpočet párovacej ceny.

Naštudovaním problematiky stereo videnia, súčasných metód výpočtu disparitnej mapy a konvolučných neurónových sietí som navrhol a implementoval riešenie pre výpočet disparitnej mapy, ktoré využíva natrénovanú neurónovú sieť pre odhad podobnosti dvoch obrázkov. Použitím siamskej architektúry siete bez poliacich vrstiev bolo možné spracovávať vstupný pár stereo obrazu ako celok. Výstupy siete sú spracovávané stereo metódou, ktorá pozostáva z agregácie a korekcie párovacej ceny a filtrácie výslednej disparitnej mapy.

Z troch testovaných sietí sa ako najúspešnejšia ukázala sieť s najmenším rozmerom vstupných obrázkov,  $9 \times 9$  pixlov, a v kombinácii s metódou spracovania disparitnej mapy dosahovala menej ako 30% chybných pixlov voči vzorovým mapám. Tým bolo dosiahnuté umiestnenie na 59. mieste z celkovo 62 porovnávaných metód v rebríčku *Middlebury Stereo Evaluation*.

Práca mi dala skúsenosť so strojovým učením, tréňovaním konvolučných neurónových sietí, optimalizáciou ich výpočtu a ich praktickom využití. Tak isto mi umožnila pochopiť problematiku počítačového stereo videnia a implementovať algoritmy zo zaujímavej oblasti spracovania obrazu.

Zámer bol splnený, a navrhnutá a implementovaná metóda generuje disparitné mapy zo vstupného páru obrazov. Pre zlepšenie kvality výstupov popísanej metódy by bolo vhodné implementovať minimalizáciu funkcie energie nad disparitnou mapou. Ďalším experimentom by mohla byť aj kombinácia viacerých neurónových sietí s rôznymi rozmermi vstupných obrázkov.

# Literatúra

- [1] Caffe deep learning framework. [Online; navštívené 11.4.2017].  
URL [caffe.berkeleyvision.org/](http://caffe.berkeleyvision.org/)
- [2] Convolutional Neural Networks (CNNs / ConvNets).  
<http://cs231n.github.io/convolutional-networks/>, [Online; navštívené 12.12.2016].
- [3] GPU-Accelerated Caffe2. [Online; navštívené 11.4.2017].  
URL [www.nvidia.com/object/caffe2.html](http://www.nvidia.com/object/caffe2.html)
- [4] TensorFlow™ - An open-source software library for Machine Intelligence. [Online; navštívené 11.4.2017].  
URL [www.tensorflow.org/](http://www.tensorflow.org/)
- [5] Torch - A SCIENTIFIC COMPUTING FRAMEWORK FOR LUAJIT. [Online; navštívené 11.4.2017].  
URL [torch.ch/](http://torch.ch/)
- [6] Bradski, G.; Kaehler, A.: *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, Inc., druhé vydání, 2013, ISBN 1449314651, 9781449314651.
- [7] Brown, M. A.; Hua, G.; Winder, S. A. J.: Discriminative Learning of Local Image Descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, ročník 33, č. 1, 2011: s. 43–57, doi:10.1109/TPAMI.2010.54.  
URL <http://dx.doi.org/10.1109/TPAMI.2010.54>
- [8] Chen, C.-S.; Lu, J.; Ma, K.-K.: *Computer Vision – ACCV 2016 Workshops*, ročník 10118. Springer International Publishing, první vydání, 2017.
- [9] Hartley, R. I.: Theory and Practice of Projective Rectification. *Int. J. Comput. Vision*, ročník 35, č. 2, 1999: s. 115–127, ISSN 0920-5691, doi:10.1023/A:1008115206617.  
URL <http://dx.doi.org/10.1023/A:1008115206617>
- [10] Hartley, R. I.; Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, druhé vydání, 2004.
- [11] Karn, U.: An Intuitive Explanation of Convolutional Neural Networks.  
<https://ujwalkarn.me/2016/08/11/intuitive-explanation-convnets/>, aug 2016, [Online; navštívené 30.12.2016].

- [12] Kriesel, D.: *A Brief Introduction to Neural Networks*. 2007, [Online; navštívené 30.12.2016].  
URL [availableathttp://www.dkriesel.com](http://www.dkriesel.com)
- [13] L., P. D.; K., M. A.: *Artificial Intelligence: Foundations of Computational Agents*. New York, NY, USA: Cambridge University Press, 2010, ISBN 0521519004, 9780521519007.
- [14] Li, L.; Zhang, S.; Yu, X.; aj.: PMSC: PatchMatch-Based Superpixel Cut for Accurate Stereo Matching. *IEEE Transactions on Circuits and Systems for Video Technology*, , č. 99, nov 2016.
- [15] Liu, T.; Fang, S.; Zhao, Y.; aj.: Implementation of Training Convolutional Neural Networks. *CoRR*, ročník abs/1506.01195, 2015.  
URL <http://arxiv.org/abs/1506.01195>
- [16] Luo, W.; Schwing, A. G.; Urtasun, R.: Efficient Deep Learning for Stereo Matching. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, dec 2016.
- [17] Marr, D.; Poggio, T.: Cooperative Computation of Stereo Disparity. Technická zpráva, Cambridge, MA, USA, 1976.
- [18] Matsugu, M.; Mori, K.; Mitari, Y.; aj.: Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, ročník 16, č. 5-6, jun 2003: s. 555–559.
- [19] Prazdny, K.: Detection of binocular disparities. *Biological Cybernetics*, ročník 52, č. 2, 1985: s. 93–99, doi:10.1007/BF00363999.  
URL <http://dx.doi.org/10.1007/BF00363999>
- [20] Qian, N.; Qian, D. N.: Binocular Disparity and the Perception of Depth. 1997.
- [21] Qian, N.; Sejnowski, T. J.: Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. In *in Proceedings of the 1988 Connectionist Models Summer*, Morgan Kaufmann, 1989, s. 435–443.
- [22] Scharstein, D.; Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, ročník 47, č. 1/2/3, apr 2002: s. 7–42.
- [23] Scharstein, D.; Szeliski, R.; Hirschmüller, H.: Middlebury Stereo Vision. [Online; navštívené 6.3.2017].  
URL [vision.middlebury.edu/stereo/](http://vision.middlebury.edu/stereo/)
- [24] Shahbazi, M.; Sohn, G.; Théau, J.; aj.: REVISITING INTRINSIC CURVES FOR EFFICIENT DENSE STEREO MATCHING. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, ročník III-3, 2016: s. 123–130, doi:10.5194/isprs-annals-III-3-123-2016.  
URL <http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/III-3/123/2016/>



- [25] Tao, H.: Image Analysis and Computer Vision. <https://classes.soe.ucsc.edu/cmpe264/Fall106/Lec12.pdf>, [Online; navštívené 15.4.2017].
- [26] Zagoruyko, S.; Komodakis, N.: Learning to Compare Image Patches via Convolutional Neural Networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2015: s. 4353 – 4361.
- [27] Zhang, C.; Li, Z.; Cheng, Y.; aj.: MeshStereo: A Global Stereo Model with Mesh Alignment Regularization for View Interpolation. *2015 IEEE International Conference on Computer Vision (ICCV)*, dec 2015.
- [28] Žbontar, J.; LeCun, Y.: Computing the stereo matching cost with a convolutional neural network. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2015.

# Príloha A

## Plagát

### Výpočet mapy disparity zo stereo obrazu s využitím konvolučnej neurónovej siete

