

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra kybernetiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Petr Z u z á n e k
Studijní program: Otevřená informatika (magisterský)
Obor: Počítačové vidění a digitální obraz
Název tématu: Detekce skoro lineárních objektů

Pokyny pro vypracování:

1. Pokračujte v bakalářské práci zadané na téma detekce skoro lineárních objektů, jako např. silnice, řeky, lesní průseky atd.
2. Navrhněte metodu pro autonomní detekci lineárních objektů z videosekvencí pořízených ze stabilizované kamerové hlavičky nesené na vrtulníku nebo bezpilotním letadle
3. Metodu implementujte v MATLAB/C++ a navažte na existující sw na pracovišti zadavatele.
4. Výslednou metodu i implementaci zdokumentujte z návrhového, programátorského i uživatelského hlediska a také experimentálně ověřte. Zdokumentujte i situace, kdy metoda přestává fungovat.

Seznam odborné literatury:

- [1] M. Šonka, V. Hlaváč, R. Boyle: Image Processing, Analysis, and Machine Vision, Thomson Learning, 2007
- [2] T. Svoboda, J. Kybic, V. Hlaváč: Image Processing, Analysis and Machine Vision, MATLAB companion, Thomson Learning, str. 73-76, 2007

Vedoucí diplomové práce: Ing. Karel Zimmermann, Ph.D.

Platnost zadání: do konce letního semestru 2012/2013


prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry




prof. Ing. Pavel Ripka, CSc.
děkan

České vysoké učení technické v Praze
Fakulta Elektrotechnická



Diplomová práce

Detekce skoro lineárních objektů

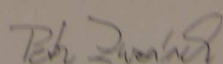
Praha, 10. května 2012

Autor: Petr Zuzánek

Prohlášení autora práce

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 10.5.2012


.....
Podpis autora práce

Poděkování

Touto cestou děkuji Ing. Karlovi Zimmermannovi, Ph.D., nejen za vedení této diplomové práce ale především za dlouhodobou spolupráci, při které mi umožnil pracovat na zajímavé úloze s reálným uplatněním. Věřím, že rady, které mi poskytl, pro mne budou zdrojem informací i v budoucí práci a jeho pohotový přístup k řešení vzniklých problémů pro mne bude motivací.

Rád bych zároveň poděkoval mé rodině za podporu při studiu.

Abstrakt

Detekování liniiových objektů není triviální úlohou kvůli jejich možnému zakrytí, či zastínění. Ačkoliv bylo publikováno mnoho přístupů k řešení této úlohy, do dnešní doby na trhu neexistuje systém řešící tento úkol, což je způsobeno především nedostatečnou obecností daných metod. V této práci je použit hybridní přístup využívající vícero rozlišení snímku, který kombinuje dobré vlastnosti metod jak pro nízké tak vysoké rozlišení. Jako forma předzpracování je použit ridge detektor v prostoru měřítek založený na diferenciální geometrii obrazové funkce. Přidanou hodnotou nalezených ridgů je získání i orientace ridge, což ubírá stupeň volnosti při prohledávání prostoru hypotéz, šetří výpočetní čas a potlačuje podíl falešně pozitivních detekcí. Dalším krokem je klasifikace získaných ridgů již na úrovni vysokého rozlišení. Pro klasifikaci je použit Gentle Adaboost klasifikátor, který se společně s haarovými popisy jeví jako dobré řešení v této úloze. Klasifikátor lze naučit na různé třídy liniiových objektů, což dělá z metody dostatečně obecný nástroj. V konečné fázi je použit linkovací algoritmus na bázi dynamického programování. Ten propojí pozitivně klasifikované ridge do výsledné liniiové struktury s ohledem na globální vlastnosti liniiového objektu.

Abstract

The detection of nearly linear objects is non trivial task due to presence of occlusions and shadows. Although several approaches have been presented in the past there is no system available on the market which generally solves this problem. This work presents the multi resolution approach which combines the advantages of both low and high resolution methods. Proposed method uses Scale-Space Ridge detector based on local differential geometry approach as a preprocessing. The detected ridges also provides the information about their orientation and thus reduce the searched space by fixing one degree of freedom. This fact saves the resources and suppress false positives. The next step consists of ridge classification on the high resolution level. The Gentle Adaboost algorithm in hand with haar-like features looks like right solution for this. In the learning phase of the classifier there is a possibility for learning of several types of linear objects which makes the proposed method general. Finally the method uses linkage algorithm based on dynamic programming and links the positive ridges together with respect to global appereance and properties of nearly-linear objects.

Obsah

1	Úvod, motivace	1
2	Formulace úlohy	3
3	Stav vědění o úloze	5
3.1	Metody extrahování silnic ze snímků s nízkým rozlišením	6
3.1.1	Metody založené na porovnávání předdefinovaných vzorů	6
3.1.2	Metody založené na dynamickém programování	7
3.1.3	Metody využívající lokální diferenciální geometrii obrazové funkce	7
3.2	Metody extrahování silnic ze snímků s vysokým rozlišením	8
3.2.1	Metody založené na porovnávání vzorů	8
3.2.2	Metody založené na učení s učitelem	9
3.2.3	Metody založené na detekci hran	10
3.2.4	Metody založené na adaptaci aktivních obrysů	10
3.3	Hybridní metody využívající více rozlišení snímků	11
4	Přehled navržené metody	13
5	Řešení	15
5.1	Ridge detektor v prostoru měřítek	15
5.1.1	Reprezentace prostoru měřítek	16
5.1.2	Formalizace světlých a tmavých ridge	17
5.2	Klasifikace ridge	19
5.2.1	Výběr popisů a jejich příznaků	20
5.2.2	Klasifikačního nástroj	21
5.3	Tvorba dominantní liniové struktury	23
5.3.1	Generování hypotézy liniového objektu	24
5.3.2	Verifikace hypotézy liniového objektu	25

5.3.3	Extrahování nejsilnější hypotézy	26
5.4	Vstupní obrazová data	28
5.5	Anotace snímků pro získání zlatého standardu	30
5.5.1	Automatická anotace silnic satelitních snímků	30
5.6	Ruční anotace silnic	31
6	Experimenty	33
6.1	Učení klasifikačního nástroje	33
6.2	Klasifikace nejsilnější hypotézy	35
6.2.1	Verifikace nejsilnější hypotézy	36
6.2.2	Výsledky klasifikace nejsilnější hypotézy podle její kvality	37
6.3	Ověření správnosti navržené metody	38
6.4	Úvaha nad použitím metody v reálném čase	39
6.5	Přesnost klasifikace pro různé počty slabých klasifikátorů	40
6.6	Případy selhání metody	41
6.6.1	Selhání na úrovni ridge detektoru	41
6.6.2	Selhání na úrovni klasifikátoru	41
6.6.3	Selhání na úrovni dynamického programování	42
7	Zhodnocení	43
7.1	Náměty pro vylepšení	44

Kapitola 1

Úvod, motivace

S postupným začleňováním počítačů do všedního života a s jejich zvyšující se výpočetní silou rostou také požadavky na ně kladené. Mezi takovéto požadavky rovněž patří zcela automatické řešení zadaných úkolů. Výjimku netvoří ani úkoly z oblasti počítačového vidění.

Tato práce je zaměřena na automatickou detekci liniových objektů z leteckých, či satelitních snímků. Pod pojmem liniový objekt jsou v terminologii používané v oblasti dálkového průzkumu země myšleny především silnice, řeky, lesní průseky, potrubní či elektrická vedení. Této zavedené terminologie se přidržím i v této práci.

Samotná detekce liniových objektů naráží na mnohé překážky. Pojem liniový objekt je poměrně abstraktní pojem, jenž v sobě skrývá velké množství entit různé barvy, velikosti, pozice, orientace i struktury. Díky jeho komplexitě lze jistou podobnost pro představu hledat např. v populárním detektoru obličejů, jenž musí být schopný detekovat obličej různé barvy pleti, s různými účesy, či barvou očí nehledě na jejich pozici kdekoliv ve snímku. Úloha detekce liniových objektů je o to těžší, že nezná (oproti zmiňovanému obličejovému detektoru) orientaci objektu a rovněž objekty nedisponují pevnou strukturou či hierarchií. Jednou z apriorních vlastností liniových objektů je jeho globální vzhled (viz. kap. 2), jenž je bohužel často znehodnocován překryvem (např. korunami stromů), stíny ap. Z tohoto pohledu tvoří úloha detekce liniových objektů v leteckých snímcích netriviální úlohu, což dokládá i fakt, že neexistuje dostupný systém na trhu řešící tuto úlohu [MH10].

Úloha je chápána jako jedna z možných aplikací nasaditelných na stabilizovanou kamerovou hlavu využívanou např. na bezpilotních letadlech. Jedním z požadavků na navádění zmíněného letadla je např. let podél takového objektu. Informace o pozici objektu je tak nezbytná pro řízení letadla, resp. kamerového systému.

Ukázka možného vzhledu liniových objektů v leteckých snímcích zobrazuje obr. 1.1.



a)



b)

Obr. 1.1: Ukázky vzhledu liniových objektů v leteckých/satelitních snímcích. Obr. a) zachycuje řeku a silnici, obr. b) zachycuje vedle silnice i železnici.

Návrh a stavba stabilizované kamerové hlavy [HHŘ⁺08] je projektem Centra strojového vnímání (CMP) Katedry kybernetiky Fakulty elektrotechnické ČVUT v Praze. Tento projekt je podpořen Ministerstvem průmyslu a obchodu České republiky v rámci projektu FR-TI1/265.

Mým úkolem je návrh modulu pro detekci lineárních objektů potencionálně použitelného v kamerové hlavě.

Kapitola 2

Formulace úlohy

Mým úkolem v této diplomové práci je navrhnout a implementovat vhodnou metodu pro autonomní detekci lineárních (liniových) objektů z leteckých, či satelitních snímků potencionálně použitelnou v kamerové hlavě (viz. kapitola 1). Liniový objekt jakožto abstraktní zástupce velké řady fyzických entit na zemském povrchu lze exaktně popsat pouze z globálního pohledu, bez zaměření se na jeho detailní vzhled. Pro všechny entity, jenž pojem liniový objekt zastupuje, jsou následující vlastnosti shodné.

1. Šířka objektu se mění plynule.
2. Směr se podél objektu mění plynule.
3. Objekt je odlišitelný od pozadí (alespoň dvojice paralelních hran).
4. Objekty jsou dlouhé, často prochází napříč celým snímkem.
5. Vzhled objektu se může měnit skokově (vlivem zastínění, překrytí, osvětlení, ap.)

V případě snímků pořízených letícím pozorovatelem nebo snímků satelitních lze za entity patřící do třídy liniových objektů považovat silnice, řeky, železniční tratě, či lesní průseky.

Na navrženou metodu jsou kladeny především následující požadavky:

1. *Autonomnost.*

Schopnost detekovat lineární objekt (viz. definice výše) zcela autonomně, tedy bez jakéhokoliv vnějšího zásahu uživatele. Motivací pro tento požadavek je skutečnost, že rychlost, s jakou se při pozorování zemského povrchu z pohybujícího pozorovatele mění scéna, může být velká. A to negativně ovlivňuje schopnost operátora korektně inicializovat poloautomatické metody.

2. *Invariance vůči velikosti objektu.*

Detekce liniového objektu libovolné šířky. Tento požadavek je důsledkem existence velkého množství liniových objektů různé šířky, které je možné pozorovat. Měnění se šířka objektu může být rovněž ovlivněna přiblížením kamerového systému, či změnou letové hladiny pozorovatele.

3. *Invariance vůči pozici a orientaci objektu.*

Pro možnost detekce libovolné silnice bez ohledu na její orientaci a pozici ve snímku.

4. *Reprezentace liniového objektu jako celku.*

Tedy s počátečním a koncovým bodem a spojitou cestou mezi těmito dvěma vrcholy. Nejedná se o úlohu diskrétní klasifikace obrazových bodů na objekt/pozadí. Pro navigační systém a případné určení směru pohybu pozorovatele je důležité identifikovat pozici celého objektu.

5. *Reálný čas*

Pro uplatnění metody na stabilizované kamerové hlavě je nutná včasná odezva systému pro řízení polohy kamerového systému příp. bezpilotního letadla samotného. Úkolem je navrhnout takovou metodu, jež je schopná tohoto požadavku dosáhnout.

Kapitola 3

Stav vědění o úloze

Metody pro extrahování lineárních objektů z digitálních, ať už leteckých nebo satelitních snímků nacházejí uplatnění ve dvou rozdílných třídách aplikací. První třídou, do které spadají i aplikace pro stabilizovanou kamerovou hlavu, jsou aplikace pro automatickou navigaci nebo dohled. Jejich úkolem je zpravidla sledovat konkrétní lineární objekt. Druhou třídu pak tvoří aplikace využívané v oblasti získávání dat pro geografický informační systém (GIS) a aktualizaci jejich databází. Cílem těchto aplikací je extrahovat všechny lineární objekty ve snímku.

Mezi pilotní dělení existujících metod patří rozdělení dle použitého rozlišení vstupních snímků, či dle míry automatizace samotné metody. I když úroveň rozlišení je z části dána hardwarovými možnostmi zařízení, jeho volbu je nutné zvážit s ohledem na charakter metody a cíle, jenž má dosáhnout. Se zvyšujícím rozlišením roste i dosažitelná přesnost detekce, naopak rychlost nalezení řešení s rostoucím rozlišením klesá. Podobné protipříklady platí i pro úroveň automatizace metody. Poloautomatické metody, jimž je apriorně předána znalost o objektu při její inicializaci operátorem, bývají často méně výpočetně náročné. Ovšem jak již bylo řečeno vyžadují vstup operátora. Oproti tomu autonomní metody pracují zcela nezávisle bez zásahu vnějších vlivů. Tuto kladnou vlastnost však mnohdy sráží jejich složitost, či delší doba zpracování.

Tato kapitola poskytuje přehled pilotních myšlenek a metod využitelných pro detekci silnic, jakožto dominantního zástupce liniových objektů, dělených dle výše zmíněných kategorií.

3.1 Metody extrahování silnic ze snímků s nízkým rozlišením

Nízké rozlišení potlačuje malé struktury, což znemožňuje získání detailního popisu objektů. Konkrétně na silnicích pak dochází k potlačení rysů, jenž ji charakterizují, především pak silničního značení, automobilů, atd. Proto v takových snímcích silnice vystupují pouze jako homogenní liniové struktury, které nelze bezpečně rozpoznat od jiných liniových struktur.

Tato abstrakce zároveň umožňuje jednotný náhled na všechny entity liniového objektu, což může být i výhodou.

Charakteristiky liniového objektu (délka, konstantní šířka, atd.) mohou být porušeny při zastínění nebo překrytí jiným objektem. Nízké rozlišení však neumožňuje díky nedostatečnému množství detailů tyto situace odlišit od falešně pozitivních detekcí.

V obecném případě pro nalezení liniové struktury v nízkém rozlišení postačí jednodušší algoritmus s malými výpočetními nároky.

3.1.1 Metody založené na porovnávání předdefinovaných vzorů

Mezi tyto metody patří jedna z průkopnických technik popsaná v [BT76]. Jedná se o automatickou metodu používající snímky s nízkým rozlišením. Toto rozlišení omezuje schopnost detekce pouze na hlavní silnice (podle [BT76] tři a více pruhové).

Algoritmus v prvním kroku aplikuje prahování, čímž vymaskuje všechny body splňující předpokládanou intenzitu silnice. Následně jsou části masky porovnávány s jedním z 52 předdefinovaných vzorů. Ty části, které se shodují s jedním ze vzorů, jsou následně spojovány do delších segmentů. Při spojování jsou používána pravidla a omezení tak, aby byly generovány přímočaré struktury s postupně měnící se orientací. Posledním krokem je zúžení nalezených segmentů na nominální šířku jeden obrazový bod a také zamítnutí krátkých segmentů.

Tento přístup předpokládá silnice jako dlouhé liniové objekty s postupně měnící se orientací. Výsledky metody jsou náchylné na volbu prahu aplikovaného při prahování, který by selhal pokud by se intenzita silnice ve snímku měnila.

3.1.2 Metody založené na dynamickém programování

Mezi zástupce této kategorie patří i automatická metoda pro detekování hlavních silnic z leteckých snímků [BC96] navržená Barzoharem a Cooperem v roce 1996. Metoda pracuje nad geometricko-stochastickým modelem silnic zohledňující střed a šířku silnice i její fotometrické vlastnosti. V prvním kroku je vstupní snímek rozdělen na čtvercové oblasti o stejné velikosti. Každá čtvercová oblast je porovnávána, zda odpovídá předdefinovanému modelu. Modelem zde jsou čtvercové výřezy odlišných segmentů silnic (přímé segmenty, zatáčky). Pokud se oblast shoduje s modelem, je konkrétní část snímku prohlášena za kandidáta. Mohou nastat případy, kdy je v oblasti více vhodných kandidátů. To signalizuje křižovatku.

V posledním kroku je pomocí dynamického programování extrahována výsledná silnice, která je složena ze získaných kandidátů. Pokud algoritmus dynamického programování prochází přes křižovatku, pak pokračuje v extrakci v jednom směru a po naplnění ukončujících podmínek se vrátí zpět na křižovatku a extrahuje novou větev.

3.1.3 Metody využívající lokální diferenciální geometrii obrazové funkce

Zcela odlišný přístup, než používají metody výše, je popsán v [Ste96]. Pomocí diferenciální geometrie obrazové funkce je pro každý bod odhadnut směr případné silnice. Orientaci silnice určují vlastní vektory Hessovy matice, která vznikla z konvoluce obrazové funkce a druhé derivace gaussovského jádra. Pro body ležící na silnici (obecně ve středu liniového objektu) je nutnou podmínkou, aby měly nulový gradient a vysoké zakřivení ve směru kolmém na směr silnice. Tímto popisem jsou maskovány takové obrazové body, které lokálně splňují vlastnost být součástí liniové struktury. Následně je nutné aplikovat linkovací algoritmus na vymaskované body pro získání liniové struktury.

Zmíněná konvoluce vede k vyhlazení obrazové funkce a potlačení detailů menších než je variance gaussovského jádra. S tím je spojena nutnost volit varianci gaussovského jádra s ohledem na šířku extrahované silnice.

3.2 Metody extrahování silnic ze snímků s vysokým rozlišením

Podobně jako používání snímků s nízkým rozlišením má i použití vysokého rozlišení své výhody a nevýhody.

Vysoké rozlišení umožňuje detailní analýzu zachycené scény. V některých případech to může být i na škodu, neboť přespříliš mnoho informací o objektu může typologicky stejné objekty klasifikovat odlišně (na silnicích se může ať změnou osvětlení či použitým materiálem měnit barva povrchu, liší se dopravní značení, příp. hustota provozu, atd.). Na druhou stranu díky detailnímu pohledu na scénu jsou snáze rozlišitelné typologicky odlišné objekty, což často při nižším rozlišení nelze. Vhodným zaměřením se pouze na určitou podmnožinu detailů, které lze z takových snímků extrahovat, je možné dosáhnout vyšší přesnosti detekce oproti metodám v nízkém rozlišení.

3.2.1 Metody založené na porovnávání vzorů

V poloautomatických metodách je v inicializační části získán reprezentativní vzor silnice doplněný o orientaci silnice (např. v [PK01] je orientace silnice získána z dominantní orientace hran ve vzoru). Následuje sledování vzoru v definovaných směrech. Pro postihnutí zatáček a zakřivení silnice jsou při sledování používány i předdefinované rotace vzoru, nejen jejich posun.

Poloautomatická metoda založená na porovnávání vzorů byla navržena v mé bakalářské práci [Zuz10] rovněž za účelem potencionálního použití na kamerové hlavě. Její inicializace je prováděna výběrem reprezentativního vzoru a oblasti zájmu, ve které se objekt nachází. V oblasti zájmu je měřena její podobnost se získaným vzorem. Následně je nad hodnotami podobností v oblasti zájmu spuštěno dynamické programování hledající cestu nejvíce podobnou vzoru. Navržená metoda měla pro aplikaci řadu omezení týkajících se především vlastností objektu (pozice, orientace, výrazný kontrast od okolí).

Nevýhodami těchto metod je množství parametrů, které je nutné nastavit, což znevýhodňuje jejich použití. Jedná se o parametry sledování (rychlost aktualizace vzoru, volba funkce podobnosti určující podobnost vzoru s jinou částí obrazu, atd.) a parametry inicializační (velikost reprezentativního vzoru, který nelze zvolit globálně ale musí být volen s respektem k šířce extrahované silnice, atd.).

Oproti tomu díky přesnému popisu reprezentativního vzoru, který je doplněn o orientaci objektu, se poměrně rapidně sníží celkový prostor hypotéz k vyhledávání. Toto je

obecně velká výhoda poloautomatických metod.

3.2.2 Metody založené na učení s učitelem

Při využití metody založené na porovnávání vzorů v automatické metodě, musel by být výše popsany koncept sledování změněn na koncept klasifikace. Úkolem v takovéto úloze pak je rozhodnout o každém bodě snímku, zda náleží silnici či pozadí pomocí klasifikátoru naučeného z reprezentativního vzorku dat s návěstím.

Tento koncept je použit např. v [MH10], kde je pro každý obrazový bod predikováno pomocí naučené neuronové sítě, zda bod náleží silnici nebo pozadí. Pro učení neuronové sítě je použito velké množství trénovacích dat, které bylo možné jednoduše generovat pomocí existujících anotací velkých datových sad satelitních snímků měst (přibližně 50 km^2). Ze zlatého standardu byla vygenerována zaručeně pravdivá silniční mapa udávající jak moc daný pixel náleží silnici. Snahou predikce je pro každý obrazový bod generovat patřičné návěstí do silniční mapy. Jsou zde zavedeny dva typy výřezů. Kontextový výřez je čtvercový RGB výřez kolem konkrétní pozice v obrázku, použitý pro učení/predikci. Tento výřez obsahuje vedle objektu zájmu (silnice) rovněž i blízký kontext. Na druhou stranu bezkontextový výřez je výsledkem predikce (při učení použit jako návěstí) kontextového výřezu a obsahuje výstupy klasifikace pouze té části kontextového výřezu, která není kontextem (ten není třeba klasifikovat). Kontextový výřez je tedy zákonitě větší (okno $64 \times 64 \text{ px}$) oproti bezkontextovému (okno $16 \times 16 \text{ px}$).

Učícím vzorkem je pak dvojice kontextový výřez vstupního obrázku a bezkontextový výřez silniční mapy (oba centrovány na stejné pozici). Pro snížení dimensionalit je použita PCA projekce vstupních dat (kontextových výřezů). Ke zlepšení kvality predikce neuronové sítě je při učení použita informovaná inicializace vah neuronové sítě (učení z dat bez učitele).

Pro získání prediktivního modelu invariantního vůči orientaci objektu zájmu je každý výřez před zahrnutím do učící procedury natočen o náhodný úhel.

V takto predikované silniční mapě jsou mnohdy silnice přerušené a místy naopak tvoří tzv. skvrny (jsou širší než by měly). K odstranění těchto artefaktů a získání vyšší predikované přesnosti je aplikováno vyhlazování silniční sítě. Toto je řešeno opět neuronovou sítí, jejíž vstupem je kontextový výřez predikované silniční mapy a výstupem bezkontextový výřez vyhlazené silniční mapy. Díky rozdílným velikostem výřezu je možné pomocí predikce zaplnit přerušovaná místa v silniční síti, stejně jako vyhladit vzniklé skvrny.

Nevýhodou metod založených na učení s učitelem je již zmíněná nutnost velkého

množství trénovacích dat. Stejně tak je nutná znalost (odhad) šířky silnice, od níž se následně odvíjí volba velikosti výřezů (kontextový i bezkontextový). Klasifikátor tak je naučen pouze na objektech určité šířky.

3.2.3 Metody založené na detekci hran

U poloautomatických metod je proces inicializace prováděný operátorem obdobný jako u metod založených na porovnávání vzorů. Operátorem je nutné zadat počáteční bod. Směr postupu ze startovního bodu může být zadán operátorem. Algoritmus uvedený v [SCR02], který extrahuje silnice, dokáže směr postupu vypočítat z dominantní orientace hran v jeho blízkém okolí. Dvojice dominantních hran jsou prohlášeny za kraje silnice a použity pro výpočet šířky objektu.

Algoritmus uvedený v [SCR02] postupuje v určeném směru. Na základě odhadu ceny zaplacené při přechodu na danou pozici se algoritmus může vydat levnější cestou úkrokem vlevo či vpravo (kolmo na směr postupu). Cena zohledňuje jak změnu intenzit pixelů tak směr postupu (snahou je postupovat co nejvíce zpříma a po oblastech se stejnou intenzitou pixelů). Pro nejlevnější nalezenou pozici je přepočítána šířka silnice a určena dominantní orientace pro další expandování. V případě prudkého nárůstu šířky silnice se jedná o křižovatku. V takovém okamžiku algoritmus expanduje pouze do jedné z větví. Po úplném prohledání této větve se algoritmus vrátí zpět na křižovatku a dělá totéž v nové větvi.

Zatímco výše popsaná poloautomatická metoda využívá pouze znalost hran, automatické metody ji používají v kombinaci s jinými technikami pro zkvalitnění výsledků. Tak tomu je i v práci [BLJ09] prezentující automatickou metodu pro extrahování silnic ze satelitních snímků měst. Extrakce je prováděna pomocí segmentace obrazu. Kombinace výsledků segmentace a hranového detektoru následně poskytuje kvalitnější výsledky než použití pouze jedné techniky.

3.2.4 Metody založené na adaptaci aktivních obrysů

Aktivní obrysy neboli hadi se používají pro extrahování liniových objektů. Kompletní teorie aktivních obrysů je popsána v [Lap97]. Výhodou aktivních obrysů oproti ostatním metodám je schopnost zahrnutí globálních geometrických vlastností hledaného objektu přímo jako formu omezení, což usměrňuje prohledávání a šetří výpočetní zdroje.

Výsledný tvar aktivního obrysu je možné ovlivnit jak geometrickými omezeními (tvar) tak fotometrickými omezeními (pozice). Geometrickými omezeními lze konstruovat poža-

navky pro spojitost, zakřivenost, či šířku aktivního obrysu. Takto definované požadavky lze interpretovat tzv. *interní silou* jakožto ukazatelem, jenž hada tlačí do tvaru, kde jsou zmíněné geometrické požadavky splněny případně minimálně porušeny.

Fotometrická omezení slouží pro definování optimální pozice hada. Při porušení fotometrických omezení vznikají tzv. *obrazové síly*, které tlačí hada za splněním fotometrických podmínek. Pro příklad, kdy jsou aktivní kontury použity pro extrahování krajů silnic, pak je fotometrickým omezením magnituda gradientu obrazové funkce. Cílem obrazové síly je aktivní obrys přesunout na pozice obrazu s vysokým kontrastem.

Samotné extrahování objektu je nutné inicializovat počátečním umístěním aktivního obrysu. S respektem k definovaným silám a inicializačnímu umístění jsou tvar a pozice aktivního obrysu iterativně optimalizovány s cílem naleznout rovnováhu mezi oběma silovými složkami. Vyrovnaní silových složek naznačuje optimalitu nalezeného řešení.

Nevýhodou je, že aktivní obrys je optimalizován lokálně, a tak nezaručuje nalezení globálně nejlepšího řešení. Jinak řečeno globálně nejlepšího řešení bude dosaženo pouze za podmínky velmi přesné inicializace. Poměrně náročnou inicializaci do jisté míry zjednoduší tzv. ZIPLOCK had [Lap97], jemuž stačí pouze definování dvou koncových bodů aktivního obrysu.

Výhodou naopak je, že díky možnosti definování zakřivenosti obrysu jsou ignorovány anomálie v obrazové funkci, a tak je možné extrahovat silnici i přes její zastíněný či zakrytý úsek.

3.3 Hybridní metody využívající více rozlišení snímků

Metody popsané v předešlých dvou kapitolách mají své výhody a nevýhody, avšak především kvůli jejich nevýhodám je nelze použít jako zcela autonomní.

U metod pro nízké rozlišení chybí dostatek detailů potřebných k verifikaci nalezené silnice, která může být zaměněna s jiným z liniových objektů.

Naproti tomu u metod s vysokým rozlišením je možné pracovat již s konkrétními detaily a charakteristikami, které se ovšem v takovýchto snímcích vyskytují i u jiných objektů. Příkladem mohou být paralelní hrany, které se vyskytují nejen na silnici, ale také na střechách budov, či na brázdách v poli.

Důsledkem v obou případech je tak generování příliš mnoha chyb při extrahování silnice.

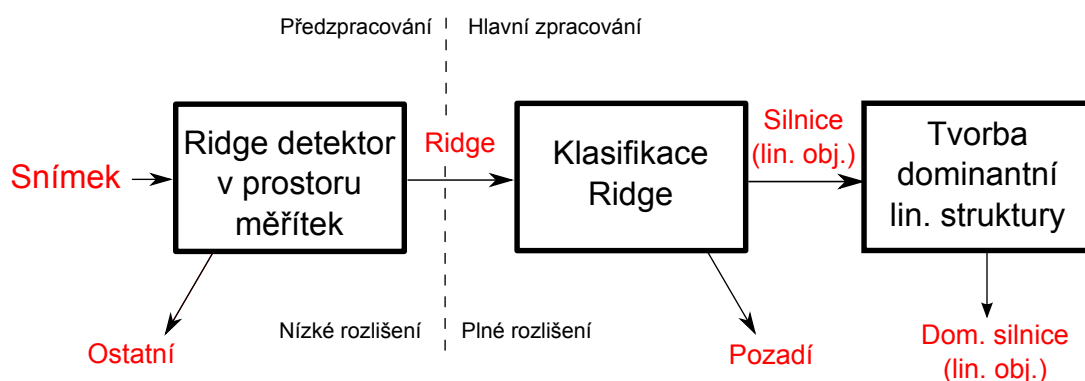
Proto se ukazuje hybridní přístup využívající vícero rozlišení jako dobrý způsob kombinující dobré a potlačující špatné vlastnosti obou výše popsaných kategorií.

Baumgartner [BSME97] využívá tento přístup k plně automatickému extrahování silnic. V prvním kroku je provedena extrakce silnic ve snímku s nízkým rozlišením. Tento snímek je přehledový obsahující mnohem více zemské plochy. S použitím této mapy zaniknou krátké falešně pozitivní detekce na střechách budov ap. Následně jsou tyto hypotézy validovány ve snímcích s vysokým rozlišením, kde jsou na pozici detekované silnice extrahovány paralelní hrany a kontrolována homogenita povrchu ve směru hran. Nalezené části silnice jsou aproximovány obdélníky. Konečné extrahování silnice probíhá v poslední etapě spojováním vzniklých obdélníků.

Nevýhodou takto popsané metody je neschopnost modelovat částečně zastíněné nebo překryté silnice obdélníky, čímž metoda selže.

Kapitola 4

Přehled navržené metody



Obr. 4.1: Schéma navrženého řešení. Červeně jsou označeny vstupy (výstupy) k jednotlivým funkčním blokům metody.

Tato část podává přehled navržené metody pro detekování lineárních objektů z videosekvencí pořízených letícím pozorovatelem a obhazuje její vhodnost s respektem k ostatním metodám diskutovaných v kap. 3 a s ohledem na požadavky kladené aplikací sepsané v kap. 2. Navržené schéma metody je zobrazeno na obr. 4.1.

S požadavkem na zcela autonomní detekci objektu koreluje výběr jedné z plně automatizovaných metod. Jako dobrou volbou se ukazuje svým popisem (viz. kap. 3) volba hybridní metody kombinující práci se snímky ve více rozlišeních. Hlavní myšlenkou je využít předzpracování snímku v nízkém rozlišení k získání liniových struktur (tedy nejen silnic) různé šířky, pozice a orientace. Následně tyto liniové struktury vhodně rozlišit dle jejich detailní analýzy, kterou umožňují snímky s velkým rozlišením.

K detekci liniových struktur je použita metoda využívající diferenciální geometrii obrazové funkce (podobné té uvedené v kap. 3) zvaná ridge detektor (Lindeberg [Lin96]). Pro možnost detekce liniových struktur libovolné šířky je žádoucí vybudovat tzv. prostor

měřitek (Lindeberg [Lin94]). Prostor měřítek je reprezentovaný vícero snímky s různým rozlišením. V tomto případě jsou různá rozlišení nahrazena originálním snímkem vyhlazeným gaussovským jádrem o dané varianci. Zvětšující se variance gaussovského jádra má za následek postupné potlačování struktur menších než daná variance, což je ekvivalentem ke snižování rozlišení originálního snímku.

Ridge detektor v prostoru měřítek poskytuje v každém bodě obrazové funkce informaci, zda bod na dané pozici může být součástí liniové struktury, a pokud ano, pak rovněž poskytuje informaci o jeho orientaci a šířce. Získaná orientace je přidanou hodnotou ridge detektoru, která ubírá stupeň volnosti prohledávaného prostoru, což je jeho velkou výhodou. Lokální vlastnosti ridge ho předurčují býti součástí podlouhlého objektu. Ridge detektor je tedy schopný detekovat téměř všechny liniové objekty (případy selhání jsou diskutovány v sekci 6.6). Ovšem hypotézu, které ridge náleží hledané liniové struktury, je nutné verifikovat vůči vzhledu hledaného objektu a globálním vlastnostem liniového objektu v hlavních krocích metody. Detailní popis ridge detektoru společně s definicí prostoru měřítek je popsán v sekci 5.1.

Hlavní zpracování probíhá ve vysokém rozlišení a skládá se ze dvou dílčích kroků. V prvním z nich je úkolem klasifikovat jednotlivé části liniové struktury (tedy výstup ridge detektoru). Klasifikační nástroj získá znalost o objektu zájmu učením z dat pomocí učitele. Tedy z předem anotovaných příkladů, jejichž návěstí tj. příslušnost ke třídě objektu je zaručeně pravdivé. Pro klasifikaci je použit binární klasifikátor naučený rozpoznávat specifikovaný objekt zájmu od pozadí. Díky tomuto přístupu je možné naučit klasifikátor rozpoznávat libovolný objekt zájmu, což přesně odpovídá požadavkům na obecný detektor liniových objektů. Popis učení klasifikátoru a výběr vhodných popisů je diskutován v sekci 5.2.

Závěrečným úkolem je s využitím globálních vlastností liniového objektu (viz. kap. 2) extrahovat dominantní liniový objekt vhodným propojením pozitivně klasifikovaných ridgů. Z důvodů možného překrytí nebo zastínění vozovky však nelze očekávat přítomnost pozitivních ridgů podél celého objektu, a tak zde není vhodné použít shlukovací popř. hladové algoritmy. Za tímto účelem je v této linkovací části použito dynamické programování, které je schopné najít globálně optimální řešení i v případě výše zmíněných komplikací. Popis extrahování nejsilnější liniové struktury popisuje sekci 5.3

Kapitola 5

Řešení

Tato část je nejprve zaměřena na podrobný popis jednotlivých funkčních bloků navržené metody (viz. obr. 4.1).

V dalších podkapitolách následuje popis dalších úkonů, které jsou s vývojem metody spojeny. Jedná se především o popis získávání vstupních obrazových dat a generování zaručeně pravdivé anotace. Dále kapitola popisuje generování množiny příkladů pro učení klasifikačního nástroje a výběr diskriminativních příznaků těchto příkladů pro dosažení nejlepší možné predikce.

5.1 Ridge detektor v prostoru měřítek

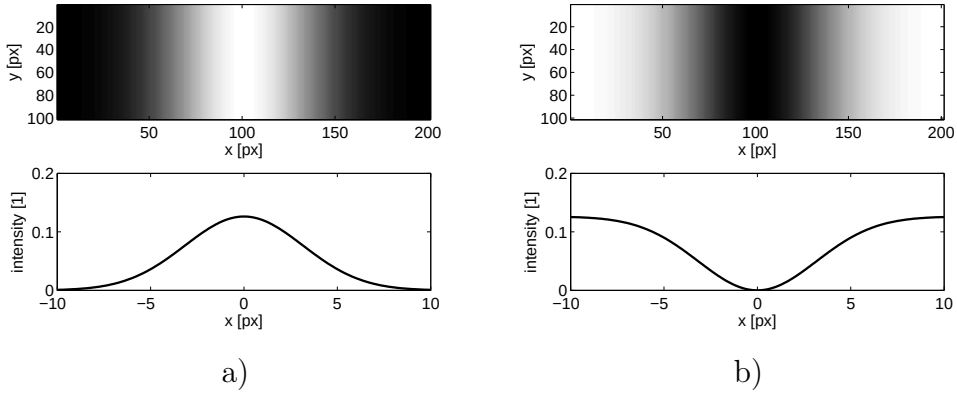
Ridge detektor je jednou z aplikací využívající lokálně diferenciální geometrii k popisu hledaného objektu. V počítačovém vidění lze ridge chápat jako horský hřeben skutečného světa. Ridge je hledán v každém bodě obrazové funkce na základě pouze lokálních pozorování. Což lze přirovnat k postavení člověka v přírodě, který se má pouze na základě lokálních pozorování (úkroky vpřed, vzad a do stran) rozhodnout, zda stojí na hřebenu, či nikoliv. Jedno z jeho uvažování by mohlo být následující:

1. Zjistit směr největšího zakřivení zemského povrchu.
2. Změnit svoji orientaci tak, aby tento směr byl paralelní s upaženými rukama.
3. Úkrokem do stran zjistit, zda se v těchto místech jeho pozice nezvýšila.
4. Pokud se krokem vpřed a krokem vzad jeho výšková pozice výrazně nezměnila prohlásit pozici za hřeben.

5. Směr pokračování hřebenu je ve směru jeho pohledu.

Výše diskutovaná úvaha popisuje pozici na hřebenu, což v korespondenci s obrazovou funkcí reprezentuje světlý ridge. Tmavý ridge reprezentuje pozici v údolí, čímž se adekvátně také změní výše popsáný způsob rozhodování. Myšlenka ale zůstane stejná.

Obr. 5.1 znázorňuje reprezentativní hřeben a údolí v obrazové funkci spolu s jejich intenzitním profilem. Směr hřebenu i směr údolí jsou ve směru osy y .



Obr. 5.1: Ukázka reprezentace hřebenu a) a údolí b) v obrazové funkci společně s jejich intenzitním profilem (dole).

Mezi prvními byl Ridge detektor popsán v přístupu Haralicka [Har83], jenž zavedl pojmy světlý (tmavý) ridge jako bod, který v dominantním směru hlavního zakřivení nabývá maxima (minima). Ve smyslu lokální diferenciální geometrie k tomuto popisu stačí zavést směrové derivace prvního a druhého řádu.

5.1.1 Reprezentace prostoru měřítek

Dříve než bude formalizován ridge detektor, je žádoucí zavést prostor měřítek, ve kterém on sám bude pracovat. Jak je uvedeno v [Lin96], velikost měřítka je úměrná velikosti variance t gaussovského jádra $g : \mathcal{R}^2 \times \mathcal{R}_+ \rightarrow \mathcal{R}$

$$g(\mathbf{x}; t) = \frac{1}{2\pi t} \exp^{-\left(\frac{\mathbf{x}^\top \mathbf{x}}{2t}\right)}, \quad (5.1)$$

Originální obraz $f : \mathcal{R}^2 \rightarrow \mathcal{R}$ pak je v daném měřítku t převeden do prostoru měřítek pomocí zobrazení $L : \mathcal{R}^2 \times \mathcal{R}_+ \rightarrow \mathcal{R}$, které je definované předpisem

$$L(\mathbf{x}; t) = g(\mathbf{x}; t) * f(\mathbf{x}), \quad (5.2)$$

kde operace $*$ značí konvoluci.

Pomocí notace výše lze zavést derivace v prostoru měřítek

$$L_{x^\alpha y^\beta}(\cdot; t) = \partial_{x^\alpha y^\beta} L(\cdot; t) = g_{x^\alpha y^\beta}(\cdot; t) * f(\cdot), \quad (5.3)$$

zde α, β značí řád derivace.

Výstup těchto operátorů je možné použít k detekci velkého množství visuálních primitiv v obraze [Lin96], jako jsou například rohy, blogy, či diskutované ridge.

5.1.2 Formalizace světlých a tmavých ridge

Se zavedeným značením v prostoru měřítek lze formulovat světlý a tmavý ridge v obrazové funkci. Pro lepší názornost je opět uvažována analogie s chůzí po hřebenu (údolí). Zpočátku je zadefinován lokální ortonormální souřadný systém v každém obrazovém bodě (x_0, y_0) , báze vektory iniciálního souřadného systému $(x, y) = \{e_x, e_y\}$ jsou zarovnány se souřadným systémem obrazu. To koresponduje s postavením člověka v krajině a jeho orientace je v souladu se světovými stranami (např. hledí na sever). Rozhodnutí, zda osoba stojí na hřebenu (v údolí), se skládá z pozorování vícero příznaků. K pozorování je nutné nejprve nalézt směr, ve kterém je největší rozptyl výškových hodnot, nazývaný též dominantní směr hlavního zakřivení. K účelu této definice je zaveden nový ortonormální souřadný systém $(p, q) = \{e_p, e_q\}$, jenž je k tomuto směru zarovnan (a směru jemu kolmém). S pomocí Hessianovy matice druhých derivací lze vyjádřit dominantní směr hlavního zakřivení v bodě (x_0, y_0) otočený vůči standardnímu (x, y) souřadnému systému o úhel β jako

$$\cos(\beta) = \sqrt{\frac{1}{2} \left(1 + \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} + L_{yy})^2 + 4L_{xy}^2}} \right)}, \quad (5.4)$$

$$\sin(\beta) = \text{sign}(L_{xy}) \sqrt{\frac{1}{2} \left(1 - \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} + L_{yy})^2 + 4L_{xy}^2}} \right)}. \quad (5.5)$$

Potom báze vektory souř. systému (p, q) jsou ve tvaru

$$e_p = (\sin \beta, -\cos \beta), \quad e_q = (\cos \beta, \sin \beta). \quad (5.6)$$

Jak ukazuje [Har83] dominantní směr hlavního zakřivení se pro světlé a tmavé ridge otáčí o $\pi/2$ [rad]. V zavedené definici směrových derivací Lindebergem [Lin96] [Lin08] je směr hlavního zakřivení pro světlé ridge ve směru báze vektoru e_p a pro tmavé ridge ve směru e_q . V případě, že tedy osoba stojí na hřebenu musí se otočit ve směru e_q a

v případě že se nachází v údolí pak ve směru (e_p). Neboli upaženými rukama reprezentuje směr hlavního zakřivení.

Zbývá učinit lokální pozorování a rozhodnout, zda je aktuální pozice na vrcholu hřebenu, na dně údolí, či ani jedno ze dvou předešlých. Tedy zda ve směru hlavní křivosti skutečně osoba stojí na nejvyšším (nejnižším místě) a zda ve směru pokračování hřebenu (údolí) bude zakřivenost povrchu dostatečně malá (pokud by byla zakřivenost velká, osoba by se nacházela buď na kopci, v jámě nebo tzv. sedle).

Výše sepsané požadavky lze pro světlý ridge formalizovat jako

$$\begin{cases} L_p = 0, \\ L_{p^2} \leq 0, \\ |L_{p^2}| \geq |L_{q^2}|. \end{cases} \quad (5.7)$$

Rovněž analogickým způsobem pro tmavý ridge

$$\begin{cases} L_q = 0, \\ L_{q^2} \geq 0, \\ |L_{q^2}| \geq |L_{p^2}|. \end{cases} \quad (5.8)$$

Směrové derivace lze vyjádřit jako lineární kombinaci parciálních derivací, tedy

$$L_p = \partial_p = \sin(\beta)\partial_x - \cos(\beta)\partial_y, \quad (5.9)$$

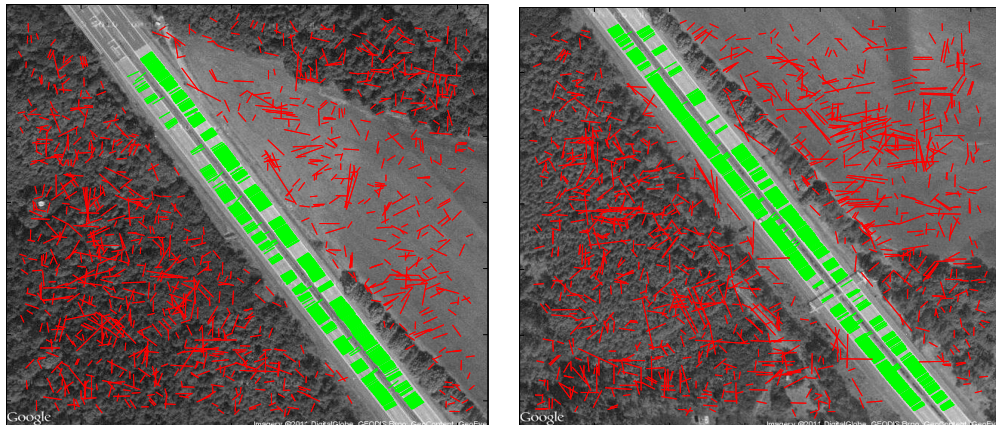
$$L_q = \partial_q = \cos(\beta)\partial_x + \sin(\beta)\partial_y. \quad (5.10)$$

Směrové druhé derivace reprezentují velikost zakřivení v jednotlivých směrech. Jejich hodnota odpovídá vlastním číslům Hessiany matice, z čehož lze dojít k následujícím definicím. Větší z nich určuje dominantní hlavní zakřivení.

$$L_{p^2} = \frac{1}{2} \left(L_{x^2} + L_{y^2} - \sqrt{(L_{x^2} - L_{y^2})^2 + 4L_{xy}^2} \right), \quad (5.11)$$

$$L_{q^2} = \frac{1}{2} \left(L_{x^2} + L_{y^2} + \sqrt{(L_{x^2} - L_{y^2})^2 + 4L_{xy}^2} \right). \quad (5.12)$$

Výhodou ridge detektoru je, že pokud nalezne v konkrétním bodě odezvu, ať už na světlý nebo tmavý ridge, poskytne zároveň cennou informaci o jeho orientaci, a díky zavedenému prostoru měřítek je rovněž získána informace o šířce ridge. Z tohoto důvodu je ridge detektor použit jako metoda předzpracování, díky které se podstatně omezí velikost prohledávaného prostoru hypotéz ridgů.



Obr. 5.2: Příklad generovaných pozitivních a negativních vzorků s respektem k anotaci. Každý ridge zobrazen úsečkou, orientace odpovídá směru hlavního zakřivení a velikost jeho šířce (ridge samotný je uprostřed úsečky). Zelené (červené) úsečky reprezentují pozitivní (negativní) vzorky (zobrazeno pouze 20% pozitivních a 2% negativních vzorků).

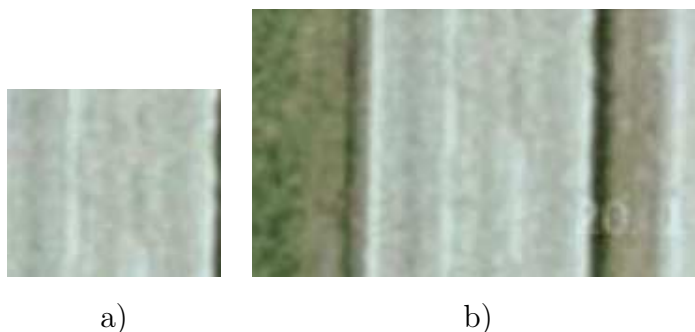
5.2 Klasifikace ridge

S naučením klasifikačního nástroje použitého pro klasifikaci nalezených ridgeů je spojeno generování reprezentativních dat, jak pozitivních tak negativních. Za tímto účelem jsou použity ručně anotované snímky (pomocí anotovacího nástroje 5.6). Pozitivní příklady zahrnují jak světlé tak tmavé ridge, jenž korespondují s anotací. To znamená, že pozitivní ridge musí ležet na anotované silnici, a jejich šířka i směr rovněž korespondují s anotací. Pozitivní a negativní ridge generované z anotovaných snímků jsou zobrazeny na obr. 5.2.

Ovšem pojmem vzorek dat se nemyslí jen konkrétní ridge ale i jeho blízké okolí, tzv. výřez. Jak je popsáno např. v [MH10], je důležité volit velikost výřezu s tím ohledem, aby byl co nejvíce zachován i kontext, ze kterého byl vzorek vyříznut. Bez příslušného kontextu pak není možné naleznout diskriminativní popisy dat bezpečně oddělující pozitivní a negativní třídy v příznakovém prostoru.

Důležitost kontextu demonstruje obr. 5.3, kde a) obsahuje výřez bez kontextu a b) obsahuje výřez s kontextem. Ze snímku a) nelze jednoznačně identifikovat typ objektu ve výřezu. Naproti tomu kontext v ukázce b) zvyšuje výpovědní hodnotu výřezu.

Díky již zmíněným vlastnostem ridge detektoru je známa šířka každého ridge, tedy šířka hypotetického objektu ridgeem reprezentovaného. Využitím této informace lze pro



Obr. 5.3: Vliv kontextu při klasifikaci. a) bez kontextový výřez, b) výřez s kontextem.

každý ridge určit velikost výřezu s ohledem na zachování kontextu.

V důsledku nutnosti zachování kontextu je každý výřez různých ridgů reprezentován odlišným počtem obrazových bodů (pouze ridge na stejném měřítku jsou reprezentovány výřezem stejných rozměrů). To může být problém při hledání reprezentativních popisů dat.

5.2.1 Výběr popisů a jejich příznaků

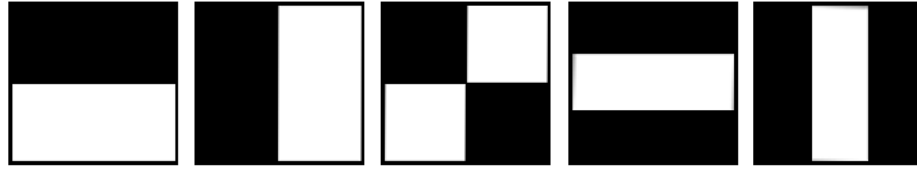
Kvalitní naučení klasifikátoru závisí mimo dat samotných také na volbě jejich popisu, podle nichž jsou data klasifikována.

Volba popisu je doménově závislý problém, tj. pro různou kategorii objektů se hodí různé popisy. V této kategorii je tedy prostor pro experimentování.

Dále se práce zaměří na klasifikaci silnic. Silnice jako člověkem vystavený objekt se vyznačuje často výrazným kontrastem vůči okolí, jeho krajnice jsou navzájem paralelní. K identifikaci také může pomoci doprovodné dopravní značení (rovněž souběžné s krajnicemi), či přítomnost aut. Samozřejmě ač je silnice kontrastní, bývá kontrastní díky svému nepřirodnímu povrchu, barva a struktura pozadí je téměř libovolná.

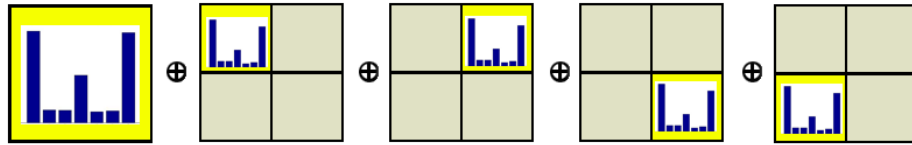
Hledány jsou tedy popisy vystihující orientaci kontrastních objektů, případně popisy pracující do jisté míry s pevným konceptem situace (osová souměrnost silnice, tj. souměrnost podél osy ridge). Takovými popisy jsou např. Haarovy popisy [VJ01] nebo hierarchický histogram orientovaných gradientů, dále HHOG (často používán pro detekci chodců [DT05], [ZZA⁺06]).

Na obr. 5.4 jsou znázorněny typy Haarových popisů použitých v této práci. Na tomto místě bych velmi rád poděkoval Ing. Karlovi Zimmermannovi, PhD. za poskytnuté zdrojové kódy pro generování velkého množství Haarových popisů těchto tvarů.



Obr. 5.4: Přehled použitých typů Haarových popisů.

Hierarchický histogram orientovaných gradientů je navržen jako konkatence pěti histogramů (každý o sedmi binech) z různých (pod)částí snímku, viz. obr. 5.5.

Obr. 5.5: Schema tvorby hierarchického histogramu orientovaných gradientů. Operace $\oplus$ značí konkatenci.

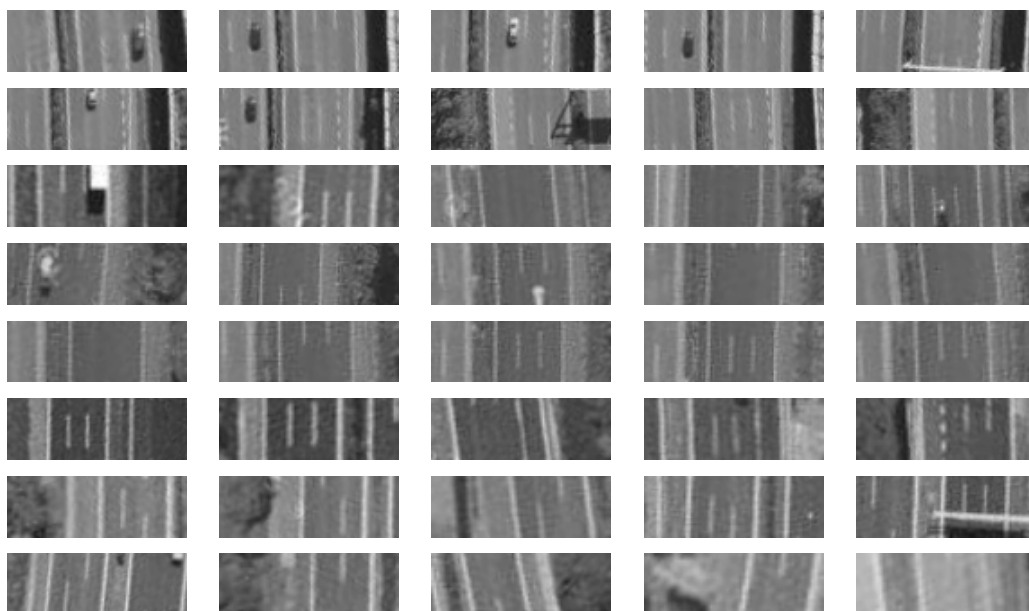
Pro získání normalizované sady dat je nutné zbavit se známé orientace ridge. U Haarových popisů je normalizace řešena rotací vzoru o známou orientaci. U HHOGů je nejprve vypočítán histogram a teprve jeho biny jsou adekvátně posunuty, což vede na rychlejší výpočet než zmíněná normalizace.

Oba výše zmíněné popisy jsou závislé na počtu pixelů, resp. na velikosti výřezu. Proto je nutná normalizace výřezu na jednotný rozměr, až poté může následovat výpočet popisů. Ukázka pozitivních příkladů je na obr. 5.6 a negativních příkladů na obr. 5.7.

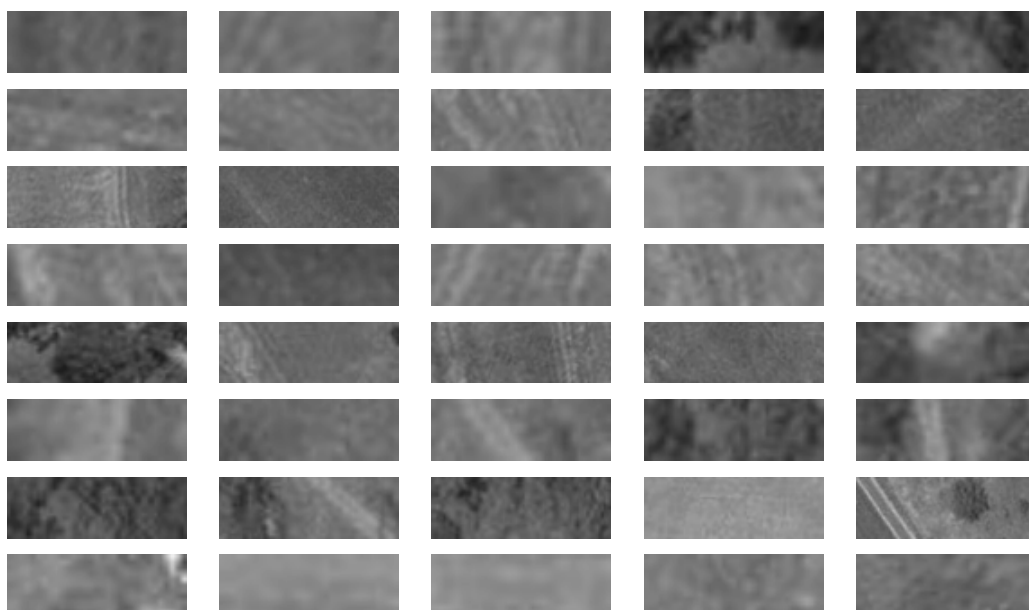
Zvolené popisy vlastně reprezentují množinu všech možných příznaků z domény daného popisu. Úkolem tak i po volbě samotného popisu zůstává, jakou minimální podmnožinu příznaků zvolit, aby tvořila dostatečně diskriminativní popis. Manuální výběr těchto příznaků je téměř neřešitelný úkol. Způsob výběru příznaků je popsán v sekci 5.2.2.

5.2.2 Klasifikačního nástroj

Jako klasifikační nástroj je zvolen boostovací algoritmus odvozený od základní verze Ada-boostu [VJ01] [Sch99], nesoucí název Gentle Adaboost (Gentleboost) [FHT00]. Koncept boostovacích algoritmů spočívá na postupném kombinování slabých klasifikátorů, ze kterých je sestaven výsledný silný klasifikátor. Slabé klasifikátory pracují pouze s hodnotami konkrétních příznaků a výběrem slabých klasifikátorů jsou tak určeny i nejvíce diskri-



Obr. 5.6: Pozitivní vzorky pro učení klasifikátoru.



Obr. 5.7: Negativní vzorky pro učení klasifikátoru.

minativní příznaky. Jediným požadavkem na slabý klasifikátor je, aby jeho klasifikační schopnost byla lepší, než je klasifikační přesnost náhodného klasifikátoru ($\epsilon = 50\%$). Díky tomu jsou slabé klasifikátory často navrhovány tak, aby jejich výpočet nebyl náročný.

Výsledný silný klasifikátor pak dokáže konkurovat jednoduchým klasifikátorům (SVM, Neuronové sítě). Silná vlastnost oproti jednoduchým klasifikátorům je schopnost automatic-

kého výběru nejvhodnějších diskriminativních příznaků. A tak jsou boostovací algoritmy, v kombinaci s Haarovými příznaky, často používaným klasifikačním nástrojem v počítačovém vidění (rozpoznávání obličejů, očí, úst, atd.).

Gentleboost používá jako slabý klasifikátor regresní funkci, která se snaží minimalizovat váženou kvadratickou chybu. Jak je ukázáno např. v [LKP03] dosahuje lepších výsledků než klasický Adaboost.

5.3 Tvorba dominantní liniové struktury

Doposud uvedené řešení se nachází ve fázi, kdy jsou extrahovány body všech liniových struktur v obraze (Ridge detektor) a u každého ridge je známo, zda může být součástí liniového objektu, který je hledán, či nikoliv (Gentleboost). Poslední nezodpovězenou otázkou tak zůstává, jakým způsobem využít pozitivní ridge k nalezení jedné dominantní liniové struktury.

Nalezení liniové struktury v množině pozitivně klasifikovaných ridgů je formulováno jako optimalizační úloha dynamického programování [SKH08]. Dynamické programování slouží pro snížení vlivu falešně pozitivních/negativních detekcí ridge detektoru. Interpretovat lze tuto úlohu také jako úlohu nalezení nejkratší souvislé cesty v trojrozměrném grafu ($2D$ prostor pozice + $1D$ prostor měřítek) tak, aby cesta propojovala co největší množství pozitivních ridgů s podobnými vlastnostmi. Podobnými vlastnostmi lze rozumět směr, pozici i šířku ridge. Pro zohlednění podobnosti ridgů je základní algoritmus dynamického programování (popsaný např. v [Zuz10]) rozšířen do prostoru měřítek a zároveň drobně modifikován. Algoritmus je modifikován ve smyslu, aby byla preferována cesta spojující ridge podobných vlastností.

Princip algoritmu je naprosto totožný s [SKH08], proto zde jsou uvedeny pouze modifikace pro rozšíření do tří dimenzí a zohlednění podobnosti.

Nechť při hledání optimální cesty algoritmus postupuje po řádcích od zdola nahoru (pro jiné směry je to zcela analogické). Cena do pixelu $Y_{i,j,s}$ (řádek i , sloupec j , měřítko s) může vést pouze z jeho přímých předchůdců $\mathcal{P}_{i,j,s}$. Pro každý pixel $X_t \in \mathcal{P}_{i,j,s}$ je definována jeho cena $F(X_t)$ (určující optimální cenu, za kterou lze dojít do bodu X_t od dolní hrany snímku). Potom optimální cesta do pixelu $Y_{i,j,s}$ vede přes:

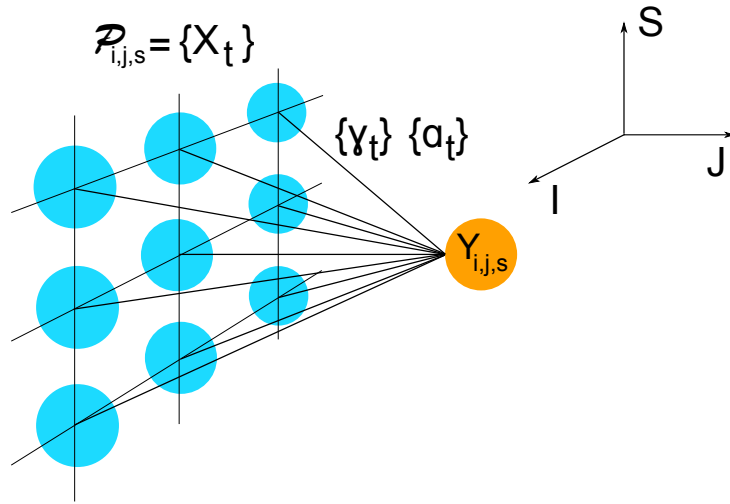
$$\arg \min_{X_t \in \mathcal{P}_{i,j,s}} (F(X_t) + \alpha_t(X_t, Y_{i,j,s})\gamma_t(X_t, Y_{i,j,s})\mathcal{H}(Y_{i,j,s})) , \quad (5.13)$$

kde $\alpha_t(\cdot, \cdot)$ $\beta_t(\cdot, \cdot)$ jsou penále za přechod mezi různými měřítky, resp. odlišnost v orientaci

bodů (nejen ridge). Funkce $\mathcal{H}(Y_{i,j,s})$ je definována předpisem

$$\mathcal{H}(Y_{i,j,s}) = \begin{cases} -\text{sign}(H(Y_{i,j,s})) & Y_{i,j,s} \text{ je ridge} , \\ 0 & \text{jinak} , \end{cases} \quad (5.14)$$

kde funkcí H se rozumí výstup klasifikátoru. Tedy funkce $\mathcal{H} : \mathcal{R}^3 \rightarrow \mathcal{R}$ pro ridge obrací znaménka klasifikovaných tříd (minimalizace), jinak ponechá nulu. To bude mít za následek, že optimální cesta povede především přes pozitivně oklasifikované ridge, poté přes body, jenž nejsou ridgi a až v krajně nevýhodném postoji bude muset projít přes negativně klasifikovaný ridge. Obr. 5.8 ukazuje entity dynamického programování použité v jednom jejím cyklu, jenž je popsán rovnicí 5.13.



Obr. 5.8: Entity grafu použitého v jedné iteraci dynamického programování (Uzly jsou obrazové body, hrany znázorňují relaci být přímým předchůdcem).

5.3.1 Generování hypotézy liniového objektu

Nástroj pro nalezení optimální cesty je tímto navržen. Zbývá určit způsob jeho použití. K dispozici totiž není globální informace o pozici ani orientaci objektu pro extrahování, navíc se jak orientace tak pozice objektu může měnit. Proto je žádoucí vytvořit dostatečně obecný způsob hledání optimální cesty. Lze apriorně předpokládat, že liniový objekt je natolik dlouhý, že ve snímku nekončí, tedy protíná libovolné dvě hrany snímku. Tím pádem je nutné použít dynamické programování takovým způsobem, aby bylo schopné najít liniovou strukturu k libovolné hraně snímku z libovolné jiné hrany. Určit tyto dvě hrany apriorně není jednoduchý a často ani jednoznačný úkol.

Problém lze přeformulovat z hledání liniové struktury mezi dvojicí hran na hledání liniových struktur z vybraného pozitivního ridge ke hranám snímku. Díky pozitivní klasifikaci ridge se předpokládá jeho náležitost k liniové struktuře a tím i hypotéza, že daným ridgem skutečně hledaný objekt prochází.

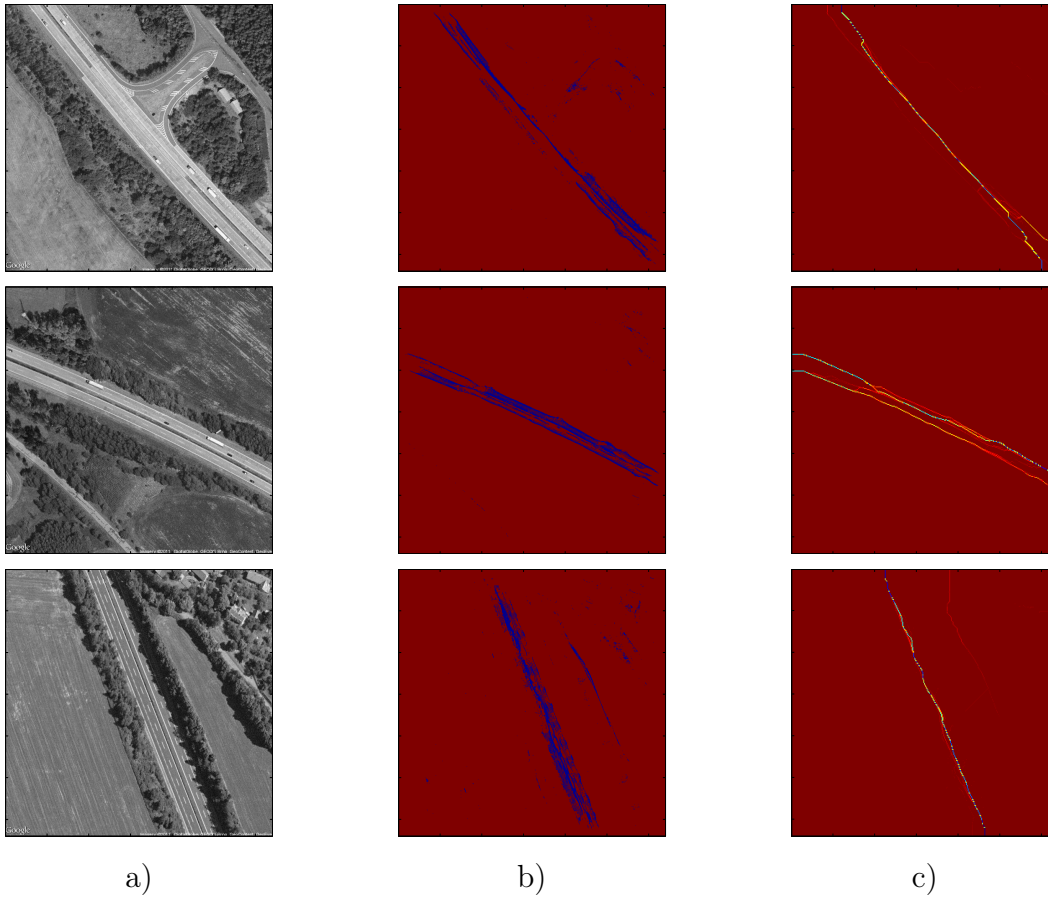
S využitím tohoto předpokladu lze použít dynamické programování k nalezení nejkratších cest z předem (náhodně) vybraných pozitivních ridgů ke všem hranám snímku. Z těchto čtyřech jsou dle stanoveného kritéria vybrány dvě nejsilnější cesty, jejichž spojením vznikne spojitá cesta od jedné hrany snímku k jiné. S nadsázkou řečeno bude tímto způsobem získána liniová struktura, jenž koresponduje s inicializačním pozitivním ridgem. Ve skutečnosti lze na nalezenou cestu nahlížet jako na jednu z hypotéz (daný pozitivní ridge může být falešně klasifikovaný), kde by se hledaná liniová struktura mohla nacházet. Je ovšem nutné tuto hypotézu verifikovat.

5.3.2 Verifikace hypotézy liniového objektu

Verifikace liniové struktury vychází z předpokladu, že hypotéza, která koresponduje s hledaným objektem (nebo jeho částí), má velkou podporu i od jiných hypotéz vygenerovaných z jiných pozitivních ridgů. Pro zachycení podpory hypotéz je použit tzv. akumulátor, do kterého jsou přičítány do patřičných pozic (body náležící dané hypotéze) hlasy za všechny vygenerované hypotézy, které danou pozicí prochází. Akumulátor, reprezentující prostor všech cest, tedy slouží k vyjádření podpory konkrétní cesty pomocí hlasování. Rovněž lze na akumulátor pohlížet jako na histogram, kde bin je reprezentován pozicí v prostoru měřítek a četnost hlasů v binu je úměrná podpoře, která daný bod předurčuje k tomu býti součástí výsledného extrahovaného objektu.

Pro měření kvality silnice je použita cena cesty získaná dynamickým programováním a normovaná svojí délkou. Mohou nastat i případy, kdy se více hypotéz liniové struktury zcela nepřekrývá, ač reprezentují stejnou liniovou strukturu. Jejich vzájemná podpora by tedy měla rovněž růst. Jednou z možností jak toto ošetřit v akumulátoru, je použít akumulátor s nižším rozlišením než má originální prostor měřítek. V této práci je použito poloviční rozlišení akumulátoru oproti prostoru měřítek.

Ukázka nejsilnějších odezev (podpory) v akumulátoru na obrazových pozicích vybraných přes prostor měřítek (tedy zobrazení $3D \rightarrow 2D$) je znázorněna na obr. 5.9.

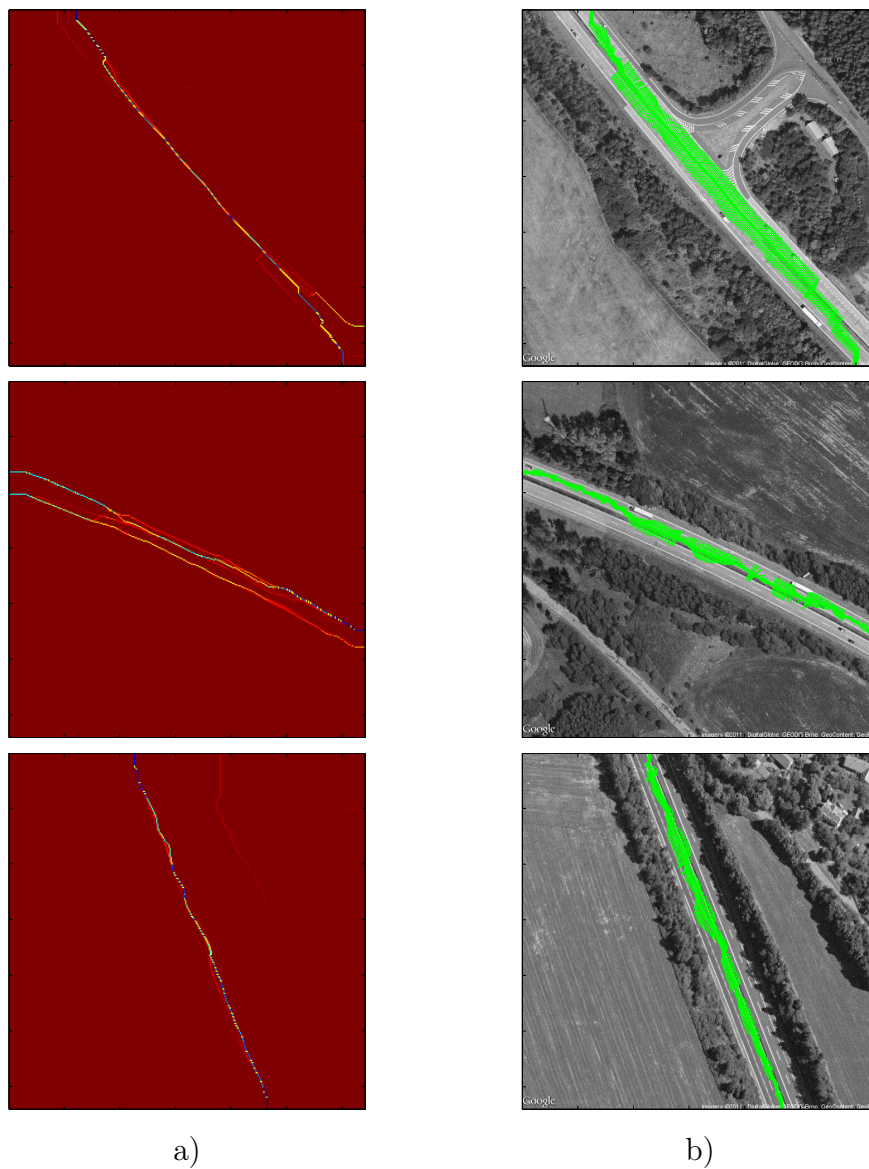


Obr. 5.9: Akumulátor pro verifikaci hypotéz. Sloupec a) originální obrázek, b) pozitivní odezvy klasifikátoru (projekce minimální odezvy funkce $\mathcal{H}(\cdot)$ do prostoru pozice, pozitivní odezvy zobrazeny modře), c) akumulátor pro 100 hypotéz. Barevně odlišeny výsledné síly hlasu (červená = 0, žlutá = -100, modrá = -250). Opět použita projekce do prostoru pozice.

5.3.3 Extrahování nejsilnější hypotézy

Ukazuje se, že vytvořený akumulátor poskytuje robustní nástroj pro verifikaci nejsilnější hypotézy i pro malé množství generovaných hypotéz liniových struktur (řádově sta). Nejsilnější hypotéza vznikne složením částí hypotéz s nejvyšší podporou v akumulátoru, opět pomocí pravidel dynamického programování. Formálně se tedy jedná o nalezení nejlevnější cesty v prostoru akumulátoru (cesta přes biny s vysokou podporou z vygenerovaných hypotéz). Nejsilnější hypotéza je generovaná analogickým způsobem popsáním výše, pouze inicializačním bodem není náhodně vygenerovaný pozitivní ridge ale bod s globálně největší podporou v akumulátoru. V tomto případě se již nemusí jednat o pozitivní ridge.

Obr. 5.10 ukazuje extrahované nejsilnější hypotézy z předem vytvořeného akumulátoru.



Obr. 5.10: Vyextrahování nejsilnější hypotézy z akumulátoru. Sloupec a) Akumulátor, barevně odlišeny výsledné síly hlasu (červená = 0, žlutá = -100, modrá = -250). Zobrazeno pomocí projekce minimálních hodnot do prostoru pozice. Obr. b) Extrahovaná dominantní hypotéza (rovněž vizualizovány informace o šířce a orientaci).

5.4 Vstupní obrazová data

V této experimentální fázi jsou namísto snímků pořízených přímo kamerovou hlavou použity snímky z Google map. Výhodou snímků z Google map je jejich velké množství, rozmanitost i možnost získání map v různém měřítku. Naproti tomu snímky pořízené z kamerové hlavy neposkytují dostatečné množství reprezentativních příkladů silnic. Mapy byly staženy v rozlišení 640×640 obrazových bodů pomocí Google API [oGAt]. Google API je jednoduše řečeno návod, jakým zkonstruovat patřičný HTML dotaz, jehož výsledkem po položení dotazu bude satelitní snímek z určené pozice (definované zeměpisnou šířkou a délkou) v definovaném měřítku popř. s dalšími definovanými vlastnostmi (rozlišení).

Struktura dotazu je následující

`http://maps.googleapis.com/maps/api/staticmap?parameters`

kde *parameters* je posloupnost příkazů specifikujících konkrétní dotaz.

Tyto příkazy mají strukturu `nazev_parametru = hodnota_parametru&`. Důležitými parametry pro tuto úlohu jsou

- *center* Pro určení zeměpisné polohy středu snímku. Udává se v zeměpisné šířce a délce.
- *zoom* Definuje úroveň přiblížení. Čím větší hodnota parametru tím blíže zemskému povrchu.
- *size* Velikost snímku. Maximální možné dosažitelné rozlišení bez prémiového účtu u Google¹ je 640×640 px.
- *maptype* Určuje druh satelitní mapy (mapa cest, satelitní mapa, hybridní mapa)
- *format* Definuje formát výstupního obrázku.
- *sensor* Požadavek Google API na jeho přítomnost. Jeho hodnota indikuje zda aplikace používá lokalizační sensor pro detekci aktuální pozice. Pro naše potřeby je hodnota nastavena na false.

Výsledný HTML dotaz tak může být např. tvaru:

`http://maps.googleapis.com/maps/api/staticmap?...`

`...center = 40.714728, -73.998672&zoom = 12&...`

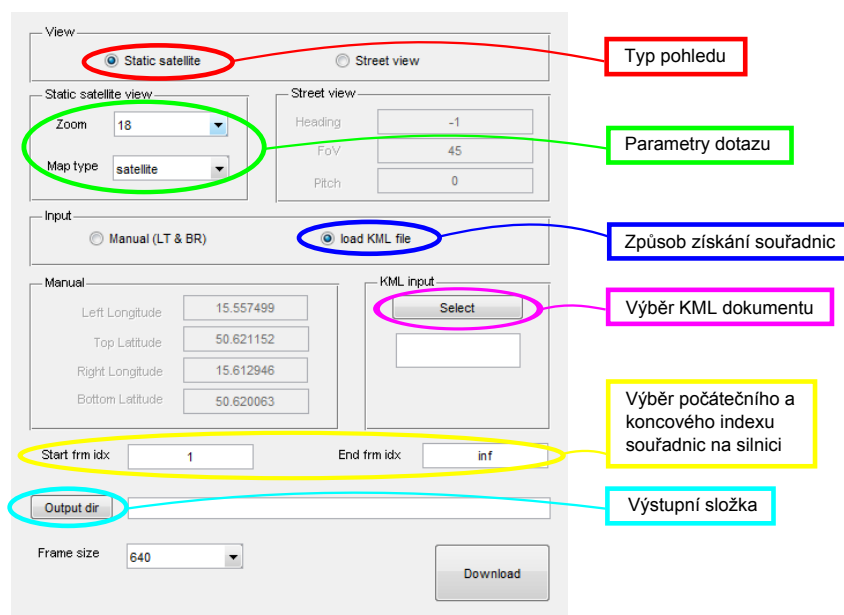
`...size = 640x640&maptype = satellite&format = jpg&sensor = false`

¹Prémiový účet umožňuje stažení snímků stejným způsobem ovšem v maximálním rozlišení 2048×2048 pixelů.

Nyní je již znám postup pro získání satelitního snímku ze známých souřadnic s různými parametry. Zbývá získat souřadnice, na kterých se nacházejí potřebná data, tj. silnice.

Pro získání souřadnic silnice je použit plánovač tras na Google mapách, kde lze naplánovat libovolnou trasu mezi různými dvěma lokalitami s možností průjezdu přes jiné lokality. Tuto trasu lze přímo uložit do `.kml` dokumentu, jenž je obdobou XML se stejnými náležitostmi. Dokument obsahuje diskretizované části naplánované trasy reprezentované v zeměpisných souřadnicích. Parsováním dokumentu tak lze získat jejich souřadnice a pomocí google API položit dotaz na získání příslušné cesty.

Pro automatizované stahování satelitních snímků podél naplánovaných tras je navržena aplikace s grafickým uživatelským rozhraním v prostředí MATLAB (dále stahovač snímků), viz. obr. 5.11. To jednoduše umožní načtení a parsování zvoleného KML dokumentu, nastavení parametrů dotazu pro Google API a definování cílové lokality pro stahované snímky ².



Obr. 5.11: Grafické uživatelské rozhraní stahovače snímků a popis funkčnosti.

Stahovač snímků má širší uplatnění³ než stahování satelitních snímků pomocí KML dokumentů, proto jsou v obr. 5.11 zvýrazněny pouze položky potřebné k tomuto účelu.

²Bez vlastnictví prémiového účtu je umožněno stahování maximálně 251 snímků denně.

³Stahování pohledů do ulic (street view) a možnost získání snímků v definovaném čtvercovém výseku

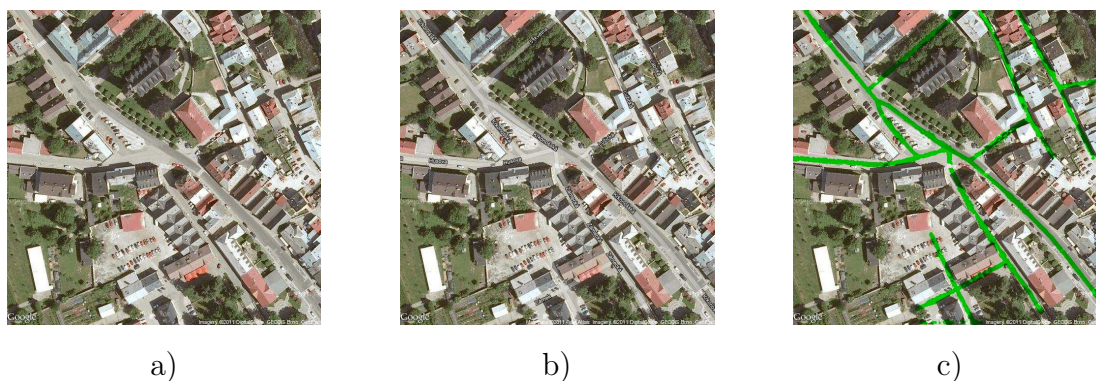
5.5 Anotace snímků pro získání zlatého standardu

Získání zlatého standardu je časově náročná úloha. Z tohoto důvodu bylo snahou tuto činnost co nejvíce automatizovat a získat tak velké množství anotovaných snímků podobně jako v [MH10].

5.5.1 Automatická anotace silnic satelitních snímků

Myšlenkou je využít kombinaci hybridního a satelitního snímku poskytovaného Google mapami k automatickému extrahování pozic silnic ve snímcích. Satelitní snímky obsahují čistý pohled na zemský povrch. U hybridních map je satelitní snímek překryt vrstvou obsahující dopravní síť, názvy významných městských budov, popisy ulic, atd. Za zlatý standard lze považovat tedy zmíněnou vrstvu hybridní mapy.

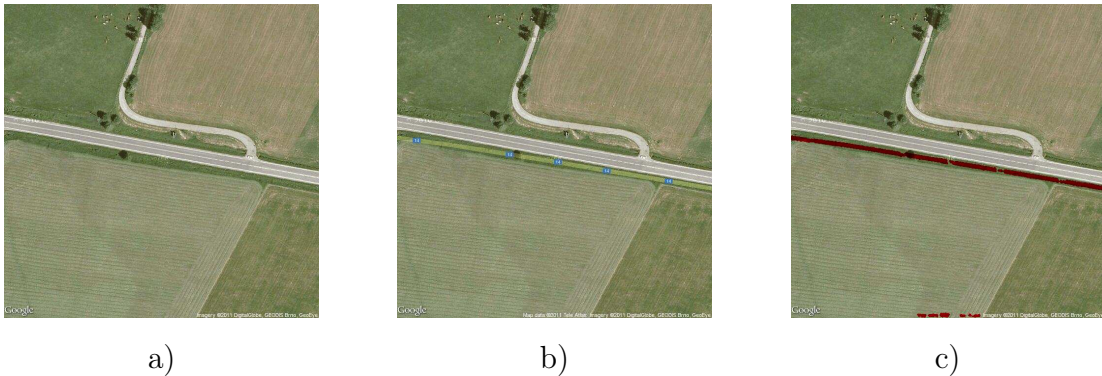
Pomocí rozdílu obou typů snímků lze extrahovat pouze překrývající vrstvu, která obsahuje i pozice silnic. Jak je již řečeno, extrahovány budou vedle silnice i další doplňující objekty (převážně nápisy a značení). Ovšem v porovnání s možnou dosažitelnou rychlostí anotace při její automatizaci je to jen malá daň. Výsledek automatické anotace silniční sítě dle navrženého postupu je zobrazen na obr. 5.12.



Obr. 5.12: Výsledky automatické anotace pomocí satelitního a hybridního snímku Google map. Obr. a) satelitní, obr. b) hybridní snímek. Obr. c) znázorňuje výsledek automatické anotace.

Bohužel se následně ukázalo, že generování zlatého standardu automatickou cestou, s využitím výše popsaného postupu, není zcela možné. Příčinou tohoto je především fakt, že silniční síť dokreslovaná do hybridního snímku často nekoresponduje se skutečnou pozicí silnice (viz. obr. 5.13). S automatickou anotací by tak muselo být spojeno poměrně

náročné předzpracování, kdy by se z množiny snímků musely ručně vybrat ty, jejichž popis v hybridní mapě koresponduje se skutečným.



Obr. 5.13: Výsledky chybné automatické anotace kvůli chybnému popisu hybridní mapy. Obr. a) satelitní, obr. b) hybridní snímek. Obr. c) znázorňuje špatný výsledek automatické anotace.

Další nevýhodou automatické anotace pomocí hybridní mapy je fakt, že hybridní popis silnice nijak nekoresponduje s její skutečnou šířkou, stejně jako je tímto postupem ztracen záznam o směru silnice. Výsledná anotace tak zachycuje pouze pozici silnice, ostatní její vlastnosti jsou potlačeny.

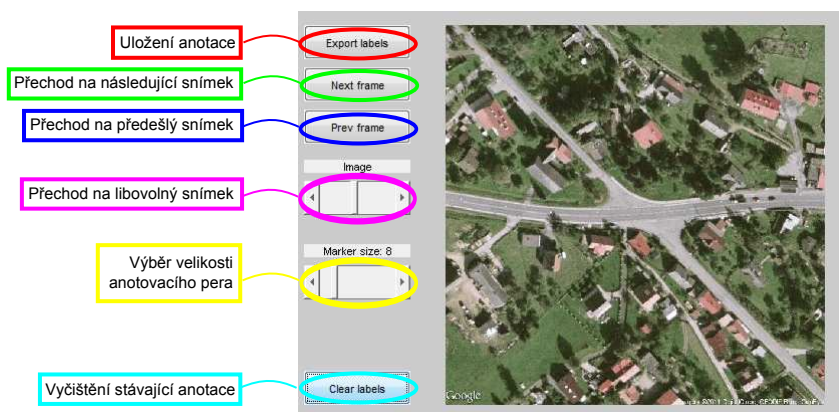
Výše popsané nevýhody nakonec vedou k závěru nepoužívat automatickou anotaci k získání zlatého standardu pro tuto práci. Ten tak bude muset být získán ručně.

5.6 Ruční anotace silnic

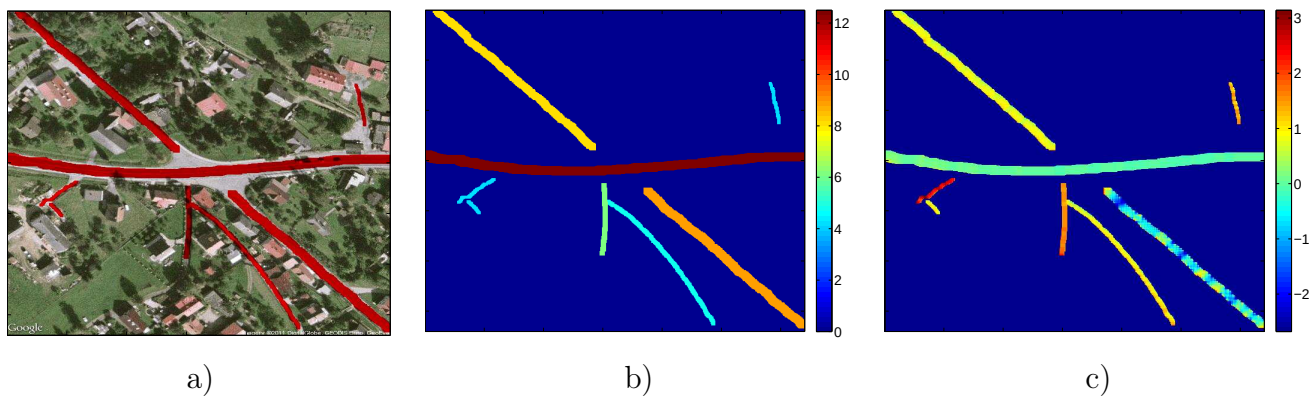
Nevýhodou výše popsaného automatického postupu, vedle již uvedených, je rovněž skutečnost, že je vhodný pouze pro snímky, u nichž existuje i hybridní popis. Takový komfort obecně není a tak je stejně nutné dané snímky anotovat ručním způsobem. Pro tento účel byla vytvořena jednoduchá aplikace s grafickým uživatelským rozhraním v prostředí MATLAB (dále anotovací nástroj), viz. obr. 5.14. Anotovací nástroj umožňuje zaznamenat pomocí anotovacího pera (ukazatel myši) pozici, přibližný směr i šířku silnice.

Směr silnice je anotován v radiánech pomocí zaznamenané pozice kurzoru v aktuálním kroku a v kroku předešlém. Označování silnice tak musí být prováděno spojitě a s ohledem na směr silnice, jenž bude zaznamenán.

Šířka silnice koresponduje s velikostí stopy anotovacího pera. Pro správnou anotaci tak musí velikost pera korespondovat s šířkou silnice.



Obr. 5.14: Grafické uživatelské rozhraní anotovacího nástroje a popis funkcí.



Obr. 5.15: Ukázka výsledku ruční anotace silnic v jednom snímku. Obr. a) anotace pozice, obr. b) anotace šířky a obr. c) anotace orientace.

Příklad anotace na satelitním snímku je uveden na obr. 5.15.

Anotovací nástroj pracuje s obrázky, jenž jsou obsahem konkrétního adresáře. Anotace pro všechny snímky jsou pro daný adresář uloženy do jednoho souboru a jejich indexace odpovídá pořadí obrázků v adresáři. Každá anotace je reprezentována jako řádká matice stejných rozměrů jako původní obrázek, viz. obr. 5.15.

Kapitola 6

Experimenty

Tato kapitola je zaměřena na ověření funkčnosti navržené metody diskutované v kapitole 5. Experimenty se zaměřují na extrahování dálnic, neboť je to kvantitativně dobře zastoupená třída liniových objektů a umožňuje generování velkého množství vzorků potřebných pro učení s učitelem. Výhodou dálnic (oproti běžným silnicím) je, že mají dostatek identifikátorů pro práci ve vysokém rozlišení, což zvyhodňuje nasazení Haarových popisů i HHOGů. Naopak jejich nevýhodou může být množství přítomných aut, které znemožňují práci ridge detektoru.

6.1 Učení klasifikačního nástroje

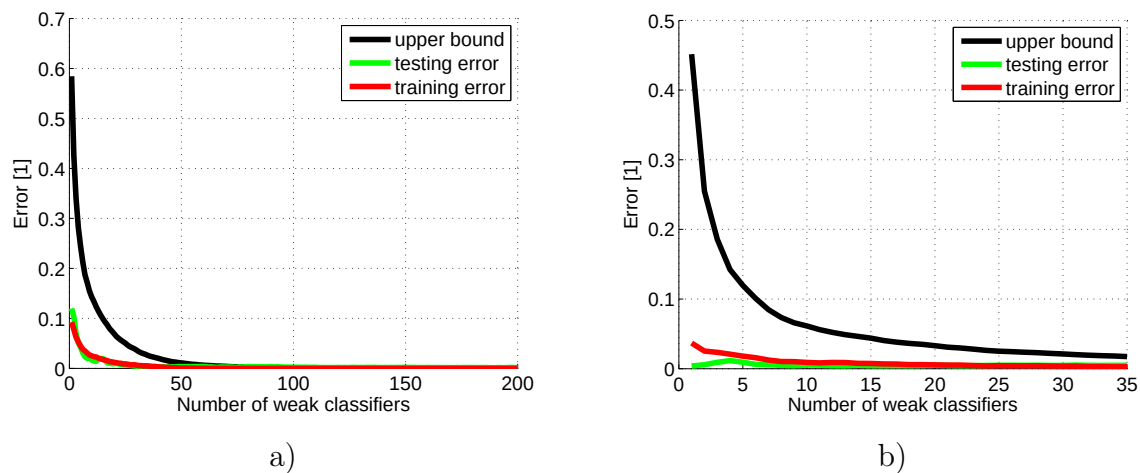
Pro získání pozitivních vzorků dálnic je použito přibližně 45 snímků dálnice, ze kterých je vygenerováno cca 6000 pozitivních vzorků a 30500 negativních vzorků. Všechny používá Gentleboost pro učení a testování. Obr. 6.4 znázorňuje učicí fáze Gentleboostu jak nad Haarovými příznaky a) tak nad HHOGy b).

K porovnání získaných klasifikačních nástrojů je použita ROC křivka falešných detekcí, viz. obr. 6.2. Pro daná data detektor nad Haarovými příznaky generuje menší poměr falešných detekcí, a proto bude použit v dalších experimentech.

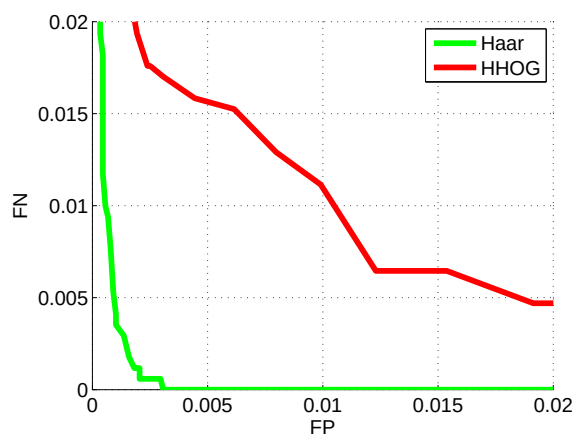
Při učení tohoto klasifikačního nástroje bylo použito 1000 příznaků pro popsání jednoho vzorku. Z nich bylo v učicí fázi vybráno 200, jejichž nalezené regresní funkce jsou nejvíce diskriminativní. Ty jsou následně používány pro predikci.

Obr. 6.3 znázorňuje tři nejvíce diskriminativní Haarovy příznaky na náhodně vybraných pozitivních vzorcích.

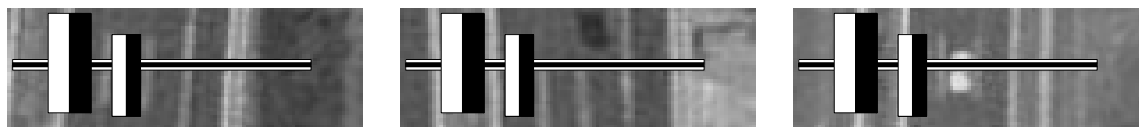
Při generování vzorků pro učení je snahou vytvořit vzájemně odlišné množiny dat. Takové aby množina pozitivních dat obsahovala pouze ridge dálnice a zároveň, aby mno-



Obr. 6.1: Typické fáze učení Gentleboostu a) pomocí Haarových příznaků, b) pomocí HHOgů.



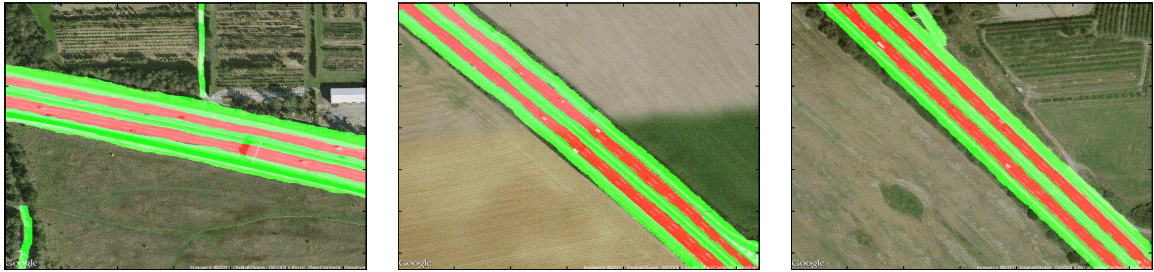
Obr. 6.2: ROC křivka falešných detekcí na testovacích datech. Parametrem ROC křivky je hodnota posunutí klasifikačního prahu.



Obr. 6.3: Tři nejvíce diskriminativní Haarovy příznaky vybrané při učení Gentleboostu.

žina negativních dat neobsahovala ridge z jiných silnic. V tomto důsledku jsou ručně anotovány dálnice jako pozitivní třída a ostatní sinice jako neutrální třída. Ridge z neut-

rální třídy nefigurují při učení ani v pozitivních ani v negativních datech. Ukázka takové anotace je zobrazena na obr. 6.4.



Obr. 6.4: Ukázka anotace pro generování trénovacích a testovacích dat.

Červená barva značí pozitivní třídu, zelená barva značí neutrální třídu. Ostatní značí negativní třídu.

Ukázka generovaných dat je na obr. 5.2. Šířka vzorků reflektuje požadavek na zachycení kontextu a je tedy volena třikrát větší než velikost ridge. Následně jsou získané vzorky normalizovány na rozlišení 100×34 obrazových bodů. Normalizované vzorky jsou zobrazeny na obr. 5.6 a 5.7.

Množina dat je pro učení rozdělena na testovací a trénovací množinu v poměru 0.7/0.3 (trénovací/testovací data). Chybu klasifikace během procesu učení obou množin dat znázorňuje obr. 6.4.

Následné experimenty probíhaly na cca 500 testovacích snímcích dálnice (cca 110 km), které byly ručně anotované (pozitivní snímky). A také na cca 400 testovacích snímcích neobsahujících žádné silnice (negativní snímky).

Prostor měřítek je popsán kompletně 50 různými měřítky v rozmezí 4 až 40 (měřítkem je směrodatná odchylka $\sigma = \sqrt{t}$ určující velikost gaussovského jádra viz. kap. 5).

U dynamického programování je přechod mezi měřítky penalizován 0.3 (jinak $\alpha_t(\cdot, \cdot) = 1$) a směrová podobnost je popsána kosinovou vzdáleností (resp. její absolutní hodnotou) mezi směry hlavního zakřivení (tedy $\gamma_t(\cdot, \cdot) = 1$ pro stejný směr, $\gamma_t(\cdot, \cdot) = 0$ pro kolmý směr).

6.2 Klasifikace nejsilnější hypotézy

Z každého testovacího snímku je extrahována nejsilnější hypotéza, společně se záznamem ukazatelem kvality. Kvalita hypotézy má rozhodující vliv pro její klasifikaci. Dynamické programování vždy nalezne nejsilnější hypotézu (pokud existují pozitivně klasifikované

ridge), ovšem její kvalita vypoví o tom, zda se jedná o pravdivě pozitivní detekci (vysoká kvalita), resp. falešně pozitivní detekci (nízká kvalita). Vhodný práh pro danou aplikaci oddělující pravdivě a falešně pozitivní detekce je možné nalézt pomocí ROC křivky. Hypotézy, které jsou pod prahem (minimalizace) jsou klasifikovány jako pozitivní a hypotézy nad prahem jako negativní. Pro zhodnocení výsledku je nutné výsledné rozhodnutí verifikovat s pomocí anotací testovacích snímků.

6.2.1 Verifikace nejsilnější hypotézy

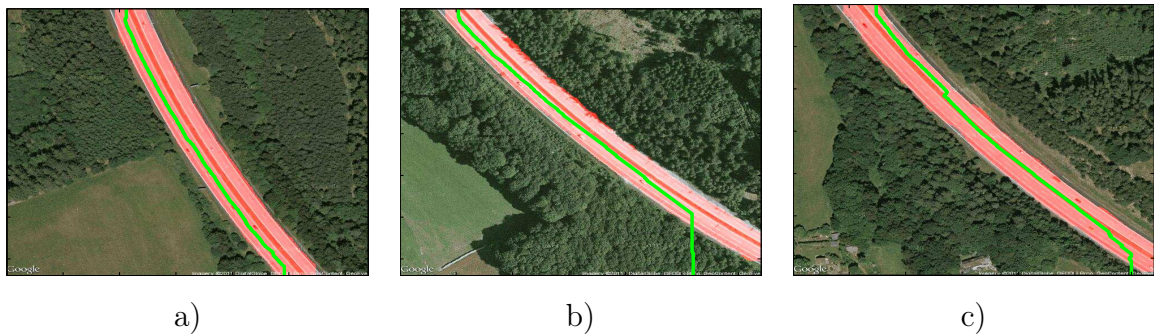
U pozitivních snímků je extrahovaná nejsilnější hypotéza porovnávána s anotací, jenž koresponduje s dominantní dálnicí ve snímku. Na této verifikační úrovni není brána v potaz šířka dálnice, či její průběžná orientace. Jediným hodnotícím ukazatelem je, zda celá extrahovaná hypotéza leží uvnitř anotace, viz. obr. 6.5, tedy zda odpovídá svojí pozicí. V podstatě mohou nastat tyto případy:

1. Nejsilnější hypotéza je klasifikována pozitivně a leží uvnitř anotace. Toto pozorování je verifikováno jako pravdivě pozitivní detekce.
2. Nejsilnější hypotéza je klasifikována pozitivně a leží z části nebo zcela mimo anotaci. Tedy snímek postrádá korektní detekci a zároveň obsahuje falešnou detekci. Toto pozorování je verifikováno zároveň jako falešně negativní i falešně pozitivní detekce.
3. Nejsilnější hypotéza je klasifikována negativně a ať leží uvnitř nebo mimo anotaci je toto pozorování verifikováno jako falešně negativní detekce.
4. Nejsilnější hypotéza leží uvnitř anotace, pouze na okrajích snímků ujíždí mimo ni. Toto je důsledkem toho, že ridge vzniklé u okrajů snímku nelze klasifikovat, neboť nelze pořídit jejich výřez odpovídajících rozměrů. Pokud tedy nejsilnější hypotéza není shodná s anotací pouze u okrajů snímků, lze rovněž toto pozorování verifikovat jako pravdivě pozitivní.

V těchto experimentech byla experimentálně nastavena toleranční vzdálenost 30 obrazových bodů od okrajů snímku.

Pro negativní snímky, neobsahující žádné silnice, lze vyvodit analogická pozorování:

1. Nejsilnější hypotéza je klasifikována pozitivně, pak lze toto pozorování verifikovat jako falešně pozitivní detekce.



Obr. 6.5: Verifikace extrahované nejsilnější hypotézy (zelená) porovnáním s anotací (červená). a) Nejsilnější hypotéza leží celá uvnitř anotace. b) Nejsilnější hypotéza leží z části uvnitř anotace. c) Nejsilnější hypotéza leží uvnitř anotace, pouze na kraji snímku z ní vyběhá (dáno neinformovaností - viz. výše).

2. Nejsilnější hypotéza je klasifikována negativně, což je verifikováno jako pravdivě negativní detekce.

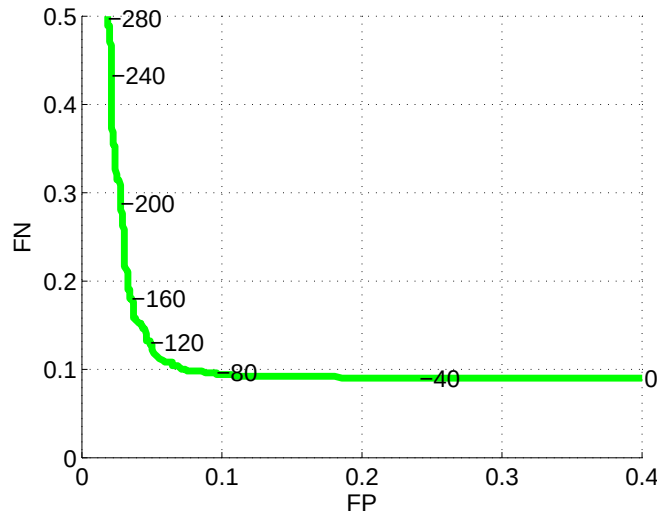
Shrnutí výše popsaných pozorování shrnuje tabulka 6.1, kde (P/N) značí pozitivní/negativní klasifikaci a (T/F) pravdivé resp. falešné rozhodnutí.

Tab. 6.1: Přehled rozhodnutí při klasifikaci a verifikaci.

Testovací snímek	P	P	P	P	N	N
Klasifikace	P	P	N	N	P	N
Verifikace	P	N	N	P	N	N
Rozhodnutí	TP	FP,FN	FN	FN	FT	TN

6.2.2 Výsledky klasifikace nejsilnější hypotézy podle její kvality

Kvalita silnice je přímo úměrná šanci, aby hypotéza byla prohlášena za silnici. Její kvalita roste se snižující se hodnotou (důsledek minimalizační úlohy dynamického programování). ROC křivka na obr. 6.6 znázorňuje poměr falešných detekcí v závislosti na hodnotě prahu, který odděluje nalezené nejsilnější hypotézy dle jejich kvality na pozitivní či negativní detekce.

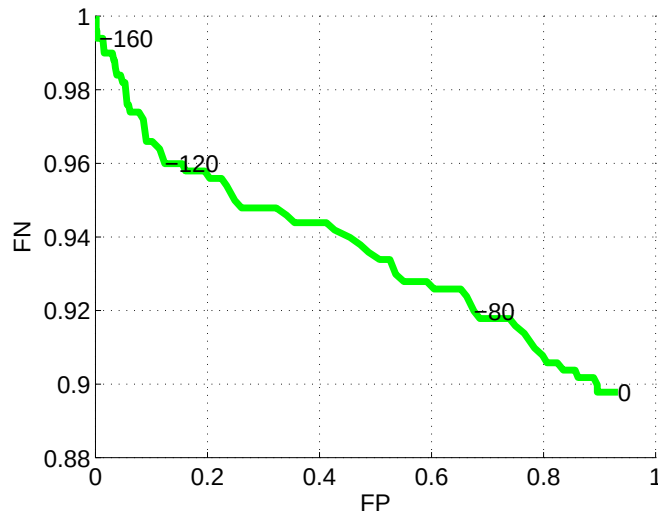


Obr. 6.6: ROC křivka falešných detekcí v závislosti na prahování kvality hypotézy. Na ROC křivce jsou pro orientaci uvedeny hodnoty prahu v daném bodě. Falešně negativní detekce (FN) se mohou vyskytovat v pozitivních snímcích testovací sady v případě, že nalezená hypotéza neodpovídá anotaci a/nebo její kvalita je nadprahová. Falešně pozitivní detekce (FP) mohou vznikat jak na pozitivních (leží mimo anotaci a jejich kvalita je pod prahem) tak negativních (kvalita pod prahem) snímcích testovací sady.

6.3 Ověření správnosti navržené metody

V této části je zhodnocena schopnost správné klasifikace jednotlivých kroků metody. Určujícím faktorem pro správné nalezení dominantní dálnice (korespondující s nejsilnější hypotézou) je klasifikace ridgeů. Samotný ridge detektor společně s dynamickým programováním neposkytují dostatečně silný nástroj řešící tuto úlohu. Dokladem jsou ROC křivky falešných detekcí dominantních objektů nalezených s použitím klasifikátoru (obr. 6.6) a bez klasifikátoru (obr. 6.7). Popř. si lze ve schéma 4.1 představit klasifikátor, jenž pro libovolný vstup generuje pozitivní návěští.

Pouze úvahou lze zhodnotit opačný případ, tedy takový, kde při hledání lineárního objektu není využit ridge detektor. Bez využití diferenciální geometrie je nutné pro získání rovnocenných výsledků, jenž odpovídají odezvám ridge detektoru, důkladně prohledat celý prostor hypotéz. Tento prostor má o jeden stupeň volnosti více, neboť není známa vedle pozice a měřítka ani orientace. Ridge detektor tak redukuje velikost prohledávaného prostoru ale zároveň nepřehlédne téměř žádné liniové objekty (případy selhání viz.



Obr. 6.7: ROC křivka falešných detekcí dominantních objektů získaných bez použití klasifikátoru ridgů. Parametrem ROC křivky je prahová hodnota kvality nalezené nejsilnější hypotézy, obdobně jako je tomu na obr. 6.6.

kap. 6.6).

Efektivní a přitom důkladné prohledávání celého stavového prostoru aplikuje metoda klouzavého okénka popsaná v [VJ01] [VJ02]. Metoda klouzavého okénka bývá použita např. pro detekci obličejů. Její nevýhodou, oproti ridge detektoru, je zmíněný stupeň volnosti za neznámou orientaci navíc, což navyšuje prohledávaný prostor a vede k většímu počtu falešně pozitivních odezev.

6.4 Úvaha nad použitím metody v reálném čase

Ridge detektor tedy neprohledává prostor orientací, čímž sníží velikost prohledávaného prostoru, a omezí generování falešně pozitivních odezev. Jeho nasazení se tedy jeví jako výpočetně přijatelnější. Toto tvrzení podporuje myšlenku použití navržené metody pro detekci v reálném čase, neboť dnešní detektory obličejů fungují v reálném čase a jejich nízké nároky dokládá nasazení na, v dnešní době běžných, fotoaparátech. Nutností zůstává vybudování prostoru měřítek, což je ale dobře paralelizovatelná procedura.

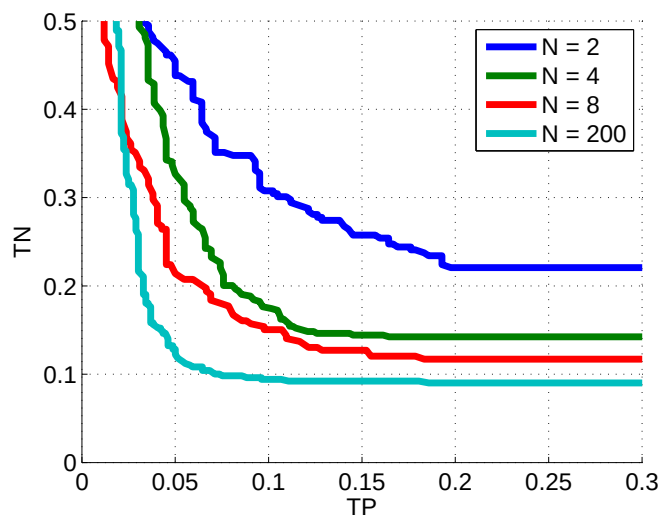
Dalším faktorem ovlivňujícím rychlost metody je množství příkladů, jenž je nutné klasifikovat. Toto je možné řešit obdobně jako při detekci obličejů tzv. kaskádou slabých klasifikátorů [VJ02]. Kaskáda elegantně řeší úsporu výpočetních zdrojů včasným odha-

lením negativních vzorků, jichž je mnohonásobně více než pozitivních. Včasná identifikace umožní jejich vyloučení z dalšího (často i náročnějšího) procesu klasifikace. Kaskáda začíná zpravidla sadou jednoduchých klasifikátorů, na kterých je vyloučen velký díl negativních výřezů, tvořící podstatnou část klasifikovaných dat. Klasifikační práh proto bývá nastaven tak, aby minimalizoval počet falešně negativních detekcí, tedy aby zamezil nechtěnému vyloučení pozitivních dat.

Na druhou stranu lze, díky schopnostem dynamického programování, zvýšit rychlost klasifikace použitím malého počtu slabých klasifikátorů, které generují minimální poměr falešně pozitivních detekcí. Na to se zaměřuje následující experiment.

6.5 Přesnost klasifikace nejsilnější hypotézy pro různé počty slabých klasifikátorů

Cílem experimentu je ověřit schopnost správné klasifikace nejsilnější hypotézy i pro nízký počet slabých klasifikátorů. Výsledky demonstruje obr. 6.8 kde jsou vyneseny ROC křivky falešných detekcí nejsilnější hypotézy. Pro její získání byl použit různý počet N slabých klasifikátorů (vždy bylo použito N prvních slabých klasifikátorů získaných ve fázi učení).



Obr. 6.8: ROC křivky falešných detekcí nejsilnější hypotézy. Parametrem ROC křivek je prahová hodnota kvality nalezené nejsilnější hypotézy, analogicky jako je tomu na obr. 6.6.

6.6 Případy selhání metody

Mohou nastat případy, kdy extrahování objektu reprezentované hlavní hypotézou selže. Příčinou je selhání jednoho, či více funkčních bloků metody (viz. obr. 4.1). Selhání každého bloku typicky spouští odlišná příčina. Z lokálních selhání je v některých situacích možné vyváznout bez výrazné újmy, především díky schopnostem dynamického programování. Rozbor jednotlivých selhání je uveden níže.

6.6.1 Selhání na úrovni ridge detektoru

Tímto se rozumí případy kdy na liniových strukturách nejsou vygenerovány ridge. Příčin pro toto selhání může být více a leckdy jim nelze předcházet.

Jednou z příčin je přítomnost šumu ve snímku, díky kterému vzniknou na liniové struktuře znehodnocující artefakty. Tomuto lze částečně předcházet snížením míry komprese. Další neovlivnitelnou příčinou je např. částečné zakrytí objektu jiným objektem (auta na silnici), či jeho zastínění (korunami stromů). Rovněž neovlivnitelnou událostí je kompozice liniového objektu vůči svému okolí. Pokud intenzitní profil silnice a jeho kontextu tvoří schodovitý profil, pak tato struktura nepůsobí jako ridge, ale jako hrana a tudíž nelze ridge detektorem rozpoznat. Tato situace na použitých snímcích nenastává, díky přítomnosti krajnic případně uniformnímu vzhledu pozadí.

6.6.2 Selhání na úrovni klasifikátoru

Neboli generování falešně pozitivních, či falešně negativních detekcí.

Selhání v této části může být způsobeno nedostatečně reprezentativní množinou příkladů použitou při učení klasifikátoru. Vzhled objektu se rovněž výrazně mění se změnou světelných podmínek, při které mohou zanikat i výrazné charakteristiky důležité pro správnou identifikaci objektu. Všem těmto komplikacím lze v ideálním případě předejít výběrem reprezentativní množiny příkladů zastupující skutečně každou instanci objektu. Získání takové množiny je ve skutečnosti téměř neřešitelný úkol. Avšak při využití sekvence snímků a s pomocí vhodného sledovacího algoritmu je možné používaný klasifikátor postupně přeučovat, čímž bude schopný dobře reagovat na případnou změnu vzhledu objektu.

6.6.3 Selhání na úrovni dynamického programování

Selhání tohoto bloku často generuje falešně pozitivní/negativní nejsilnější hypotézy liniové struktury.

Dynamické programování pracuje pouze s hodnotami, které poskytnou předešlé funkční bloky metody. Vždy extrahuje nejsilnější cestu (pokud není akumulátor prázdný kvůli neexistujícím pozitivním ridgům), ta ovšem ne vždy musí být v souladu se sémantikou snímku. Pokud například snímek obsahuje rozsáhlou silniční síť, nelze jednoznačně určit dominantní silnici, a tak nemusí korespondovat s očekávaným výsledkem. I v tomto případě by pomohlo využití sekvence snímků společně se sledovacím algoritmem.

Kapitola 7

Zhodnocení

Tato diplomová práce byla zaměřena na návrh metody pro autonomní detekování liniových objektů ve videosekvencích pořízených letícím pozorovatelem. Jako nejvhodnější volbou se jeví metoda spadající do kategorie metod, které pracují s vícero rozlišeními snímku.

Nízké rozlišení je použito při fázi předzpracování pro získání odezev ridge detektoru, které určí nejen o pozici liniových objektů, ale rovněž o jejich šířce a orientaci. Pro získání šířky objektu je žádoucí vybudovat prostor měřítek, orientace objektu je vedlejším produktem lokální diferenciální geometrie, kterým je ridge popsán. Ridge získané abstraktním popisem liniového objektu v nízkém rozlišení mohou náležet různým entitám ve skutečném světě. Jejich následná identifikace probíhá zcela ve vysokém rozlišení.

Pro jejich identifikaci je použit klasifikační nástroj Gentle Adaboost, který lze vhodně naučit pro detekování patřičné třídy liniových objektů. Klasifikátoru jsou předkládány ridge, resp. jejich obdélníkové výřezy. Šířka výřezů je volena s ohledem na šířku ridge tak, aby výřez obsahoval i blízký kontext. Výška výřezu je dopočítána s ohledem na zachování poměru stran po následné normalizaci na jednotný rozměr.

Nad pozitivně klasifikovanými ridgi je v poslední fázi metody spuštěn linkovací algoritmus založený na dynamickém programování pro získání dominantní liniové struktury. K extrahování dominantní liniové struktury patřičného typu liniového objektu je použit akumulátor. Ten počítá podporu každého bodu, která vyjadřuje jeho dílčí šanci stát se součástí této dominantní struktury. Příspěvky v akumulátoru odpovídají vygenerovaným cestám nalezeným dynamickým programováním, které je spouštěno opakovaně z různých pozitivně klasifikovaných ridgů. Tyto cesty jsou generovány v prostoru měřítek a jejich snahou je během hledání cesty projít co nejvíce pozitivně klasifikovaných ridgů. Výsledná nejsilnější liniová struktura je hledána v prostoru akumulátoru tak, aby procházela body

s vysokou podporou.

V experimentální části byla ověřena robustnost navržené metody na dálnicích. Klasifikátor Gentleboost byl naučen přibližně z 10 km dálnic a použit pro detekci dálnice na přibližně 110 km validačních snímků dálnice. Na validačních datech byla dominantní liniová struktura ve většině případů extrahována správně. Experimenty rovněž ukázaly velkou robustnost navrženého akumulátoru, který i pro malý počet akumulovaných cest (řádově sta, což odpovídá pouze jednotkám procent z celkového počtu pozitivně klasifikovaných ridgů), zachytí správnou liniovou strukturu. Na validačních datech obsahujících cca 500 pozitivních a 400 negativních snímků bylo docíleno úspěšné klasifikace až 92%, při daném způsobu verifikace (viz. kap. 6).

Ačkoliv je klasifikace ridgů výpočetně nejvíce náročnou operací, ukázalo se, že má zásadní vliv na přesnost extrahování dominantní liniové struktury.

7.1 Náměty pro vylepšení

Navržená metoda je pouze funkčním prototypem, jehož hlavním úkolem je demonstrovat způsob přístupu, kterým lze problém detekce liniových problémů řešit. Velký prostor pro vylepšení poskytuje časová optimalizace. Jak je psáno v kap. 6 metoda má předpoklady pro práci v reálném čase, avšak tento prototyp má k tomu zatím daleko.

Výpočetně náročnými operacemi je výstavba prostoru měřítek a rovněž i klasifikace ridgů. Obě operace, stejně jako práce s akumulátorem, lze dobře paralelizovat. Navíc u výpočtu Haarových příznaků lze namísto normalizace výřezů (velikost, orientace) škálovat a otáčet dané příznaky. Zrychlení klasifikace, resp. rychlé vyřazení negativních příkladů v brzkých fázích klasifikace, může být docíleno vystavením kaskády klasifikátorů, viz. kap. 6.

Jako formu vylepšení lze rovněž považovat schopnost detekovat nejen dominantní liniovou strukturu daného typu, ale všechny struktury. I když toto nemusí být pro navigační systémy zcela žádoucí aplikací. To především z důvodu nutné interakce s operátorem pro výběr objektu ke sledování. Každopádně pro tyto účely již není zcela vhodné použít stávajícím způsobem navržený akumulátor. Resp. by muselo dojít k předzpracování pozitivních ridgů a jejich kategorizaci (podle orientace a pozice). Tímto způsobem by bylo možné extrahovat dominantní liniové struktury v jednotlivých shlucích a tím extrahovat např. křižící se silnice, ap. Pro tento experiment již nebyl časový prostor.

Literatura

- [BC96] Meir Barzohar and David B. Cooper. Automatic Finding of Main Roads in Aerial Images by using Geometric - Stochastic Models and Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:707–721, July 1996.
- [BLJ09] D. B. L. Bong, K. C. Lai, and A. Joseph. Automatic road network recognition and extraction for urban planning. In *World Academy of Science, Engineering and Technology 53*, pages 209–215, 2009.
- [BSME97] A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein. Multi-resolution, semantic objects, and context for road extraction. In *In Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pages 140–156. Birkhauser Verlag, 1997.
- [BT76] R. Bajcsy and M. Tavakoli. Computer recognition of roads from satellite pictures. *IEEE Transactions on systems, man and cybernetics*, March 1976.
- [DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893, 2005.
- [FHT00] J. Friedman, T. Hastie, and R. Tibshirani. Additive Logistic Regression: a Statistical View of Boosting. *The Annals of Statistics*, 38(2), 2000.
- [Har83] Robert M Haralick. Ridges and valleys on digital images. *Computer Vision, Graphics, and Image Processing*, 22(1):28 – 38, 1983.
- [HHŘ+08] Zdeněk Hurák, Martin Hromčík, Martin Řezáč, Jaroslav Žoha, Pavel Krsek, Václav Hlaváč, Tomáš Pajdla, and Milan Bartoš. Inerciálně stabilizovaná kamerová základna pro bezpilotní letoun. *Automa*, -:27–31, 2008.

- [Lap97] Ivan Laptev. Road extraction based on snakes and sophisticated line extraction. In *Master's thesis, Computational Vision and Active Perception Lab (CVAP), Royal Institute of Technology*, 1997.
- [Lin94] Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.
- [Lin96] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, CVPR '96, pages 465–, Washington, DC, USA, 1996. IEEE Computer Society.
- [Lin08] Tony Lindeberg. *Wiley Encyclopedia of Computer Science and Engineering*, chapter Scale-Space, pages 2495–2504. 2008.
- [LKP03] Rainer Lienhart, Er Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *In DAGM 25th Pattern Recognition Symposium*, pages 297–304, 2003.
- [MH10] Volodymyr Mnih and Geoffrey E. Hinton. Learning to detect roads in high-resolution aerial images. In *ECCV (6)*, pages 210–223, 2010.
- [oGAt] An online Google API tutorial. Static maps api v2 developer guide.
- [PK01] Seung-Ran Park and Taejung Kim. Semi-Automatic Road Extraction Algorithm from IKONOS Images Using Template Matching. *Paper presented at the 22nd Asian Conference on Remote Sensing*, November 2001.
- [Sch99] Robert E. Schapire. A brief introduction to boosting. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.
- [SCR02] Vandana Shukla, R. Chandrakanth, and R. Ramachandran. Semi-Automatic Road Extraction Algorithm for High Resolution Images Using Path following Approach. In *Indian Conference on Computer Vision, Graphics and Image Processing*, December 2002.
- [SKH08] T. Svoboda, J. Kybic, and V. Hlavac. Image processing, analysis and machine vision: A matlab companion. 2008.

- [Ste96] Carsten Steger. Extracting curvilinear structures: A differential geometric approach. In *Proceedings of the 4th European Conference on Computer Vision-Volume I - Volume I*, ECCV '96, pages 630–641, London, UK, UK, 1996. Springer-Verlag.
- [VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 1:I-511–I-518 vol.1, 2001.
- [VJ02] Paul Viola and Michael Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2002.
- [Zuz10] Petr Zuzánek. Sledování skoro lineárních objektů v reálném čase. Research Report K333–39/10, CTU–CMP–2010–08, Department of Cybernetics, Faculty of Electrical Engineering Czech Technical University, Prague, Czech Republic, July 2010.
- [ZZA⁺06] Qiang Zhu, Qiang Zhu, Shai Avidan, Shai Avidan, Mei chen Yeh, Mei chen Yeh, Kwang ting Cheng, and Kwang ting Cheng. Fast human detection using a cascade of histograms of oriented gradients. In *In CVPR06*, pages 1491–1498, 2006.

