



České vysoké učení technické v Praze

Fakulta elektrotechnická

Katedra kybernetiky

**Vizuální odometrie
pro dynamickou rekonstrukci obrazu
Diplomová práce**

Vedoucí práce: Ing. Libor Přeučil, CSc.

Autor: Bc. Tomáš Pivoňka

květen 2018

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Pivoňka Jméno: Tomáš Osobní číslo: 420037
Fakulta/ústav: Fakulta elektrotechnická
Zadávající katedra/ústav: Katedra kybernetiky
Studijní program: Kybernetika a robotika
Studijní obor: Robotika

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Vizuální odometrie pro dynamickou rekonstrukci obrazu

Název diplomové práce anglicky:

Visual Odometry for Dynamic Image Reconstruction

Pokyny pro vypracování:

1. Seznamte se s postupy pro vizuální odometrii s důrazem na příznakově chudé scény.
2. Analyzujte možnosti a omezení využití kamerového páru pro zjištění relativního posuvu mezi dvěma po sobě jdoucími obrazy scény se vzájemným překryvem.
3. S využitím předchozího navrhnete postup pro výpočet relativního posuvu mezi dvěma snímky obrazu podvozku vozidla za jízdy. Zohledněte požadavek na zpracování v čase blízkém reálnému.
4. Navržený postup implementujte a ověřte v laboratorních/simulovaných podmínkách.
5. Zhodnoťte výslednou přesnost metody v závislosti na charakteru scény, světelných podmínkách a velikosti a druhu pohybu obrazu.

Seznam doporučené literatury:

- [1] SCARAMUZZA, Davide, Friedrich FRAUNDORFER a Wonpil YU. Visual Odometry [Tutorial]. International Journal of Control, Automation and Systems. 2011, 18(4), 80-92. DOI: 10.1109/MRA.2011.943233. ISSN 1070-9932. Dostupné také z: <http://ieeexplore.ieee.org/document/6096039/>
- [2] HARTLEY, Richard. a Andrew. ZISSERMAN. Multiple view geometry in computer vision. 2nd ed. New York: Cambridge University Press, 2003. ISBN 05-215-4051-8.
- [3] CVIŠIĆ, Igor, Josip ČESIĆ, Ivan MARKOVIĆ a Ivan PETROVIĆ. SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. Journal of Field Robotics. , -. DOI: 10.1002/rob.21762. ISSN 15564959. Dostupné také z: <http://doi.wiley.com/10.1002/rob.21762>
- [4] CHOI, Sunglok, Jaehyun PARK a Wonpil YU. Simplified epipolar geometry for real-time monocular visual odometry on roads. International Journal of Control, Automation and Systems. 2015, 13(6), 1454-1464. DOI: 10.1007/s12555-014-0157-6. ISSN 1598-6446. Dostupné také z: <http://link.springer.com/10.1007/s12555-014-0157-6>
- [5] BOULEKCHOUR, Mohammed, Nabil AOUF a Wonpil YU. Robust motion estimation using covariance intersection. 22nd Mediterranean Conference on Control and Automation. IEEE, 2014, 2014, 18(4), 1014-1019. DOI: 10.1109/MED.2014.6961507. ISBN 978-1-4799-5901-3. ISSN 1070-9932. Dostupné také z: <http://ieeexplore.ieee.org/document/6961507/>

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Libor Přeučil, CSc., inteligentní a mobilní robotika CIIRC

Jméno a pracoviště druhého(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: 10.01.2018

Termín odevzdání diplomové práce: 25.05.2018

Platnost zadání diplomové práce: 30.09.2019

Ing. Libor Přeučil, CSc.
podpis vedoucí(ho) práce

doc. Ing. Tomáš Svoboda, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

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

.....

Poděkování:

Za pomoc při vypracování diplomové práce děkuji zejména vedoucímu práce Ing. Liboru Přeučilovi, CSc. Dále bych rád poděkoval Ing. Karlu Košnarovi, Ph.D. a Mgr. Martinu Dörflerovi, kteří rovněž pracují na projektu KASSANDRA, a autoservisu BestDrive v Praze 5 za podporu při provádění experimentů.

Abstrakt:

Cílem této práce bylo navrhnout a experimentálně ověřit metodu pro vizuální odometrii určenou k měření pohybu podvozku automobilu jako součást 3D skeneru podvozků. Tento skener pracující pouze s obrazem z kamer je vyvíjen v rámci projektu KASSANDRA. Navržená metoda má sloužit pro skládání dílčích 3D modelů podvozku do celkového modelu. V úvodní části práce jsou popsány základní principy a postupy používané pro vizuální odometrii. Na základě teoretické části byla navržena metoda typu „3D do 2D“ opírající se o příznakovou vizuální odometrii. Zvolený postup je založen na silné filtraci příznaků, která umožňuje použití metody i v příznakově chudých scénách. Pro detekci a popis příznaků je používán obrazový deskriptor ORB, jehož parametry jsou získány automatickým vyhodnocením parametrů příznaků pro předpokládaná vstupní data. Metoda byla implementována, experimentálně ověřena a kvantitativně vyhodnocena na pořízené testovací množině snímků podvozku. Na závěr byla provedena analýza výpočetní náročnosti jednotlivých kroků metody.

Klíčová slova:

vizuální odometrie, příznaky, stereo kamera, počítačové vidění

Abstract:

The thesis focuses development and testing of visual odometry method serving for measurement of car undercarriage motion. The solution is foreseen to be used with a scanning system for reconstruction of the 3D shape of the vehicle undercarriage (related to project KASSANDRA). The proposed solution relies exclusively on the use of regular cameras. The method enables stitching of particular 3D undercarriage models together. The first part of this thesis describes basic principles and techniques of visual odometry. A feature-based method relying on 3D to 2D principle has been focused in this work. The core approach relies on strong filtration of visual features also allowing application in feature-sparse cases, all making-use of ORB visual features. Parameters of the method have been chosen by an automatic procedure for feature properties selection based on the typical input data. The proposed method has been implemented, tested and quantitatively evaluated with an experimental dataset of undercarriage images. The work also provides an analysis of computational intensity and resources needs for components of the proposed method.

Keywords:

visual odometry, features, stereo camera, computer vision

Obsah

1	Motivace.....	1
2	Úvod	3
2.1	Projekt KASSANDRA.....	3
3	Vizuální odometrie	5
3.1	Základní principy a dělení vizuální odometrie.....	5
3.1.1	Úvod k vizuální odometrii.....	5
3.1.2	Základní dělení metod pro vizuální odometrii	5
3.2	Geometrie kamery.....	6
3.2.1	Model perspektivní kamery.....	6
3.2.2	Stereo kamera	8
3.2.2.1	Epipolární geometrie.....	8
3.2.2.2	Výpočet pozice bodu ze souřadnic v obraze	9
3.2.3	Kalibrace kamery	10
3.3	Analýza postupů s využitím příznaků	11
3.3.1	Výpočetní vlákno příznakové vizuální odometrie	11
3.3.2	2D do 2D.....	12
3.3.2.1	Osmibodová metoda.....	12
3.3.2.2	Pětibodová metoda.....	13
3.3.2.3	Výpočet rotace a translace z esenciální matice	13
3.3.3	3D do 3D.....	14
3.3.4	3D do 2D.....	15
3.3.4.1	P6P	15
3.3.4.2	P3P	16
3.4	Detektory a deskriptory příznaků.....	18
3.4.1	Harrisův detektor	18
3.4.2	Shi-Tomasi detektor	20
3.4.3	SIFT	20
3.4.4	SURF.....	21
3.4.5	FAST detektor	22
3.4.6	BRIEF deskriptor	23
3.4.7	ORB.....	24
3.5	Přiřazování a filtrace příznaků	25
3.5.1	Přiřazování příznaků	25

3.5.1.1	FLANN	25
3.5.2	RANSAC	26
3.6	Systémy a metody pro vizuální odometrii	26
3.6.1	SOFT-SLAM	26
3.6.2	RotRocc++	27
3.6.3	Monokulární vizuální odometrie pro silniční vozidla	28
3.6.4	Zjednodušení epipolární geometrie pro kolová vozidla	30
3.6.5	Metoda pro robustní odhad pohybu	31
3.6.6	SVO	31
4	Navržené metody	33
4.1	Metoda pro automatické testování vlastností příznaků	33
4.2	Metoda pro vizuální odometrii	33
4.2.1	Detekce a přiřazování příznaků	34
4.2.2	Filtrace příznaků a triangulace	34
4.2.3	RANSAC a výpočet rotace	35
4.2.4	Výpočet translace	36
4.2.5	Určení úhlů z rotační matice	37
5	Implementace a experimenty	39
5.1	Hardware	39
5.1.1	Výpočetní jednotka	39
5.1.2	Kamera Basler acA1920-155uc	39
5.1.3	Dataset pro testování	40
5.2	Hledání optimální metody pro detekci a popis příznaků	40
5.2.1	Automatické testování a testovací množina	40
5.2.2	Harrisův detektor a BRIEF deskriptor	40
5.2.3	Shi-Tomasi detektor a BRIEF deskriptor	42
5.2.4	SIFT	42
5.2.5	SURF	43
5.2.6	FAST detektor a BRIEF deskriptor	44
5.2.7	ORB	45
5.2.8	Porovnání testovaných metod	46
5.3	Ověření navržené metody na reálných datech	46
5.3.1	Parametry výsledné metody	46
5.3.2	Výsledky testování pro světlé snímky	46

5.3.3	Výsledky testování pro tmavé snímky	48
5.3.4	Rotace.....	48
5.4	Analýza výpočetní náročnosti navržené metody	48
6	Závěr.....	51
6.1	Další rozvoj navržené metody	52
	Literatura.....	53
	Příloha A – obsah CD:.....	57

Seznam obrázků

Obr. 1	Schematické znázornění systému pro 3D skenování podvozků vyvíjeného v projektu KASSANDRA.....	1
Obr. 2	Vizualizace skeneru podvozku KERBEROS 3D od společnosti VOP CZ, s.p. Tento systém je zabudován do vozovky a snímá podvozky nad ním projíždějících automobilů. (zdroj: http://www.vop.cz/en/produkt/47-kerberos-systems.aspx).....	4
Obr. 3	Perspektivní model kamery se znázorněním souřadných systémů, zdroj [4].....	7
Obr. 4	Znázornění epipolární geometrie kamery. Body C_1, C_2 značí středy promítání kamer, e_1 a e_2 jsou epipóly, přímky l_1 a l_2 představují epipolární přímky a body u_1, u_2 znázorňují obrazy bodu X , který je určen prostorovými souřadnicemi v soustavě σ . π_1 a π_2 jsou roviny promítání. (zdroj [5], str. 114).....	8
Obr. 5	Snímky značek pro kalibraci kamery pomocí nástrojů v knihovně OpenCV (vlevo) a APRIL kalibrace (vpravo, zdroj [9]).....	11
Obr. 6	Blokový diagram vlákna pro vizuální odometrii pracující s příznaky (zdroj [2]).....	11
Obr. 7	Znázornění souřadných systémů pro řešení problému tří bodů (zdroj [14]).....	17
Obr. 8	Ackermannův podvozek a znázornění pohybu po kružnici.....	29
Obr. 9	Konstrukce s kamerami Basler pro pořízení testovacích snímků podvozku automobilu. Šipka znázorňuje směr pohybu konstrukce při snímání podvozku automobilu.	39
Obr. 10	Ukázka snímků podvozku s přiřazenými příznaky mezi 2 světlými posunutými snímky se vzdáleností podvozku 57cm, zobrazeny jsou příznaky se správným přiřazením podle RANSAC algoritmu, ze kterých bude dále určována translace.....	49
Obr. 11	Ukázka způsobu snímání podvozku automobilu pro získání experimentálních dat	51

Seznam tabulek

Tab. 1	Testované parametry pro Harrisův a Shi-Tomasi detektor a BRIEF deskriptor.....	41
Tab. 2	Optimální hodnoty parametrů a výsledky automatického testu pro Harrisův detektor a BRIEF deskriptor	41
Tab. 3	Optimální hodnoty parametrů a výsledky automatického testu pro Shi-Tomasi detektor a BRIEF deskriptor.....	42
Tab. 4	Testované parametry pro metodu SIFT.....	42
Tab. 5	Optimální hodnoty parametrů a výsledky automatického testu pro metodu SIFT.....	43
Tab. 6	Testované parametry pro metodu SURF.....	43
Tab. 7	Optimální hodnoty parametrů a výsledky automatického testu pro metodu SURF.....	43
Tab. 8	Testované parametry pro FAST detektor s BRIEF deskriptorem	44
Tab. 9	Optimální hodnoty parametrů a výsledky automatického testu pro FAST detektor s BRIEF deskriptorem	44
Tab. 10	Testované parametry pro metodu ORB	45
Tab. 11	Optimální hodnoty parametrů a výsledky automatického testu pro ORB metodu	45
Tab. 12	Parametry navržené metody pro vizuální odometrii popsané v kapitole 4.2	46
Tab. 13	Výsledky experimentálního ověření navržené metody pro světlé snímky.....	47
Tab. 14	Výsledky experimentálního ověření navržené metody pro tmavé snímky	47
Tab. 15	Testování přesnosti určení rotace	48
Tab. 16	Průměrný čas výpočtu jednotlivých kroků metody pro světlé snímky se vzdáleností podvozku 75cm.....	48

Seznam použitých zkratk:

2D/3D -dvourozměrný/trojrozměrný

SLAM - Simultaneous Localization and Mapping, úloha v mobilní robotice, kdy je určována trajektorie robotu a mapa okolního prostředí zároveň

SVD - Singular Value Decomposition, singulární rozklad matice, využívá se při řešení homogenní soustavy lineárních rovnic pomocí metody nejmenších čtverců

6DoF - Degrees of Freedom, 6 stupňů volnosti

P3P - Perspective of 3 Points, problém tří bodů

RANSAC - Random Sample Consensus, iterativní metoda pro nalezení parametrů matematického modelu z reálných dat

px – pixel, obrazový bod

Seznam často používaných symbolů:

Pozn. neplatí pro kapitolu 3.4, ve které je používáno mírně odlišné značení.

$\vec{u} = [u, v]^T$ - bod v obrazových souřadnicích

$\vec{x} = [x, y, z]^T$ - bod v souřadnicích kamery

$\vec{X} = [X, Y, Z]^T$ - bod v obecných prostorových souřadnicích

R - rotační matice

K - matice kamery

\vec{t} - vektor translace

$P = [R, \vec{t}]$ - matice spojující rotaci a translaci (v kapitole 3.3.2.2 vynásobená maticí kamery)

T - matice 4x4 spojující rotaci a translaci

E - esenciální matice

F - fundamentální matice

λ - skalár přidávající stupeň volnosti obrazovým souřadnicím v modelu perspektivní kamery

$\vec{0}$ - nulový vektor

$\vec{l} = [l_1, l_2, l_3]^T$ - vektor s koeficienty rovnice přímky ve 2D prostoru

S, V, U - výsledné matice ze singulárního rozkladu matice (SVD)

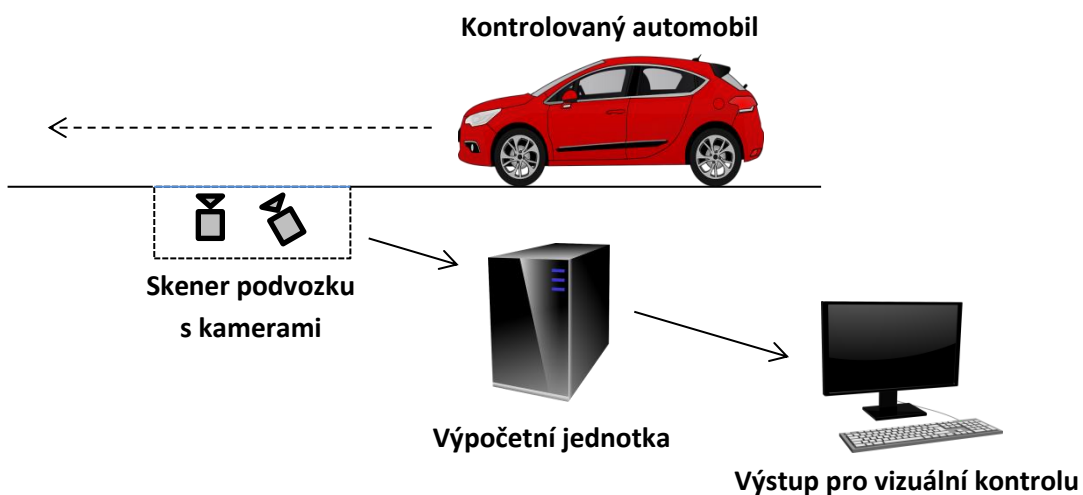
α, β, γ - úhly yaw, pitch, roll

k - index bodu z množiny

i - celočíselná diskrétní hodnota času (pořadí) snímku

1 Motivace

Vizuální odometrie slouží k určování polohy kamery nebo více kamer na základě změny v obraze způsobené pohybem kamery. Tato diplomová práce byla vypracována v rámci projektu KASSANDRA řešeného skupinou Inteligentní a mobilní robotiky CIIRC ČVUT ve spolupráci s firmou VOP CZ, s.p. Cílem celého projektu je vývoj kamerového systému pro 3D skenování podvozků vozidel. Při průjezdu automobilu nad skenerem budou vytvářeny 3D mapy jednotlivých částí podvozku, které budou spojovány do výsledného 3D modelu podvozku. Cílem této práce je návrh metody pro vizuální odometrii, která bude určovat transformace mezi jednotlivými snímky, na jejichž základě bude dále řešeno skládání dílčích 3D modelů podvozku.



Obr. 1 Schematické znázornění systému pro 3D skenování podvozků vyvíjeného v projektu KASSANDRA

1. Motivace

2 Úvod

Diplomová práce je členěna do 6 hlavních kapitol. Na stručnou kapitolu s motivací pro vypracování této práce navazuje úvodní část, ve které je kromě struktury práce podrobněji představen projekt KASSANDRA (kapitola 2.1), jehož součástí je tato práce. V této podkapitole je dále zmíněn i aktuální stav řešení projektu a z toho plynoucí nároky na tento systém.

Třetí kapitola je věnována teoretickému úvodu k vizuální odometrii. V úvodu této kapitoly (kapitola 3.1) je podrobněji vysvětlen pojem vizuální odometrie a popsáno základní dělení metod pro vizuální odometrii. V další části (kapitola 3.2) jsou představeny model perspektivní kamery, geometrie stereo kamery a metody pro kalibraci kamer. Na teoretický základ ke geometrii kamer navazuje kapitola 3.3 o příznakové vizuální odometrii s analýzou tří možných přístupů k řešení této úlohy. Pro každý přístup jsou uvedeny základní metody, ze kterých lze spočítat výsledný posun kamery. Protože je práce zaměřena na příznakovou vizuální odometrii, jsou dále představeny základní metody pro detekci příznaků v obraze a vytváření jejich deskriptorů (kapitola 3.4). Deskriptory příznaků slouží pro nalezení shodného příznaku mezi snímky. Metodám pro přiřazování příznaků mezi snímky a filtrování špatně přiřazených příznaků se věnuje podkapitola 3.5. V poslední části třetí kapitoly (kapitola 3.6) jsou uvedeny příklady již existujících systémů pro vizuální odometrii a představeny další možné přístupy k řešení úlohy vizuální odometrie. Zmíněna je například možnost využití omezené geometrie kolových vozidel pro snížení počtu stupňů volnosti pohybu vozidla.

Ve čtvrté kapitole je nejprve představena metoda pro automatické testování příznaků (kapitola 4.1). Hlavní motivací pro návrh této metody byla specifická charakteristika snímků podvozku oproti běžným prostředím, pro které je navrhována většina metod pro vizuální odometrii. Metoda je založena na testování různých kombinací parametrů pro detekci a popis příznaků a jejich ohodnocení na testovací množině snímků. Ve druhé části (kapitola 4.2) je představena navržená „3D do 2D“ metoda pro vizuální odometrii, která využívá odděleného výpočtu rotace a translace.

V páté kapitole je nejprve uveden hardware používaný při experimentech (kapitola 5.1). Dále jsou prezentovány výsledky automatického testu příznaků, na jejichž základě je zvolena metoda pro detekci a popis příznaků, která bude použita v navržené metodě pro vizuální odometrii (kapitola 5.2). Po automatickém testu následuje ověření přesnosti navržené metody na experimentálně pořízených datech (kapitola 5.3) a analýza výpočetní náročnosti jednotlivých kroků metody (kapitola 5.4).

V závěrečné kapitole (kapitola 6) jsou shrnuty výsledky dosažené v této práci a poslední část je věnována dalšímu možnému rozvoji navržených metod.

2.1 Projekt KASSANDRA

Jak již bylo zmíněno v kapitole 1, projekt KASSANDRA se zabývá návrhem skeneru podvozků automobilu pro bezpečnostní účely. Projekt navazuje na předcházející projekt KERBEROS [1] z roku 2013, jehož skener byl založen na kombinaci RGB kamery a laserového dálkoměru. Hlavní nevýhodou tohoto systému byla složitá instalace, protože spodní část senzoru se nacházela přibližně metr pod úrovní vozovky. Další motivací pro nahrazení laserového dálkoměru kamerami je zvýšení rozlišovací schopnosti, která obecně bývá u laserového dálkoměru poměrně nízká, a jejich nižší cena.

2. Úvod



Obr. 2 Vizualizace skeneru podvozku KERBEROS 3D od společnosti VOP CZ, s.p. Tento systém je zabudován do vozovky a snímá podvozky nad ním projíždějících automobilů. (zdroj: <http://www.vop.cz/en/produkt/47-kerberos-systems.aspx>)

Potřeba skenování podvozků automobilů vychází z požadavků bezpečnostních složek, které potřebují kontrolovat, zdali na podvozku nedošlo ke změnám. Nejzávažnějším důvodem je možnost umístění výbušniny, a proto jsou tyto skenery instalovány u strategicky významných objektů. Kontrola neprobíhá pouze vizuálně ostrahou objektů, ale umožňuje i porovnání se snímky stejného vozidla pořízených v minulosti. Tím je možné detekovat například změnu, ke které dojde mezi odjezdem a příjezdem vozidla zpět do objektu.

Další částí projektu KASSANDRA je aplikace metod umělé inteligence na získané snímky, aby bylo možné rozlišit typ vozidla a porovnat podvozek vůči podvozkům z vozidel stejného typu. Tím by bylo možné provádět automatickou inspekci i u vozidel, které nejsou v uložení v databázi systému.

Projekt KASSANDRA je zatím na svém počátku, kdy aktuálním cílem je vývoj prototypu a návrh jednotlivých metod. V době vypracovávání práce ještě nebyla navržena přesná metoda pro vytváření 3D modelů ani dostupné větší množství dat na testování. Taktéž zatím nejsou známe přesné parametry hardwaru. Metoda navržená v této práci bude proto spíše výchozím bodem pro návrh celého systému, než konečným řešením splňující všechny požadavky pro komerční produkt.

Zatím také nebyly stanoveny přesné požadavky na přesnost a časovou náročnost systému. Návrh metody proto bude vycházet z následujících předpokladů. Délka vozidla bude předpokládána 4,5m. Maximální posun mezi dvěma snímky bude uvažován 30cm, což při snímkovací frekvenci 50Hz umožňuje teoretickou rychlost vozidla 54km/h a 21,6km/h při 20Hz. Pro celý podvozek bude předpokládán počet snímků, pro které bude potřeba určit posun, 20 snímků. Tomuto počtu odpovídá výpočet 19 posunů mezi po sobě jdoucími snímky. Čas zpracování by neměl při testování výrazněji přesáhnout 30s. Ve výsledném systému bude délka trvání výpočtů snížena optimalizací navržené metody a použitím výkonnějšího hardwaru.

3 Vizualní odometrie

3.1 Základní principy a dělení vizualní odometrie

3.1.1 Úvod k vizualní odometrii

Vizualní odometrie je jedním ze způsobů měření pohybu, který je založen na zpracování obrazu z jedné nebo více kamer. Při pohybu dochází mezi jednotlivými snímky ke změnám, které je možné detekovat a na jejich základě určit, o kolik se kamera posunula. Termín vizualní odometrie byl poprvé použit v roce 2004 ve stejnojmenném článku [3]. Název byl zvolen na základě podobnosti s kolovou odometrií, u které je pohyb určován z enkodérů pro měření otáčení kol. Vizualní odometrii lze obecněji zařadit do skupiny úloh nazývané „structure from motion“ (určování struktury z pohybu). Jednou z prvních aplikací vizualní odometrie bylo měření trajektorie pro planetární vozidla v 80. letech 20. století, u kterých bylo potřeba vyvinout alternativu ke kolové odometrii, u níž nedochází k ovlivnění měření smýkáním kol. Zpočátku byly systémy limitovány nedostatkem výpočetního výkonu. První systémy, které byly schopné fungovat v reálném čase, začaly vznikat až po roce 2000.

V současnosti mohou být tyto systémy využívány jako dočasná alternativa k družicové navigaci, protože dokážou pracovat i v místech, kde není možné přijímat signál z družice. Obdobné systémy bývají také často postaveny na laserových dálkoměrech, které dokáží skenovat okolí ve 3D s vysokou přesností. Hlavním omezením těchto senzorů je jejich stále vysoká cena.

Obdobnou úlohou k vizualní odometrii je vizualní SLAM (V-SLAM). SLAM (simultánní lokalizace a mapování) je dnes často řešená úloha v robotice, kdy úkolem není pouze měření trajektorie, jako u vizualní odometrie, ale i vytváření mapy prostředí, která není předem známá. Řada systémů pro vizualní odometrii si však také vytváří vlastní mapu prostředí, na základě které se provádí odhad trajektorie vozidla.

Při postupném určování pozice z nově příchozích snímků dochází k akumulaci chyb určení jednotlivých posuvů. Přestože chyba mezi jednotlivými snímky může být zanedbatelná, při dlouhém měření může docházet k velkým odchylkám oproti skutečné trajektorii. Tyto odchylky lze detekovat a opravit, pokud se systém vrátí na místo, které již měřil. Tím dojde k uzavření smyčky a trajektorie může být přepočítána tak, aby i trajektorie tuto smyčku obsahovala. Detekcí smyček a přepočtem přes všechna data je dosahováno globálně konzistentního řešení. Detekce smyček a způsob výpočtu trajektorie bývá hlavním rozdílem mezi V-SLAMem a vizualní odometrií, která s detekcí smyček nepočítá a optimalizace trajektorie může probíhat pouze o několik snímků zpět, nikoliv přes celou množinu dat. Výsledná trajektorie proto bývá pouze lokálně konzistentní. Její výhodou je oproti tomu zpracovávání dat v reálném čase postupně tak, jak snímky přichází v časové řadě.

3.1.2 Základní dělení metod pro vizualní odometrii

Základní dělení metod pro vizualní odometrii vychází z počtu použitých kamer. Metody pracující s jednou kamerou jsou nazývány monokulární. Jejich hlavní nevýhodou je, že nelze jednoznačně určit měřítko vektoru translace. Pro určení měřítka lze využít informace z jiných senzorů nebo známých rozměrů snímaných předmětů. Další skupinou jsou stereo metody, které používají dvě kamery a dokáží jednoznačně určit polohu snímaného bodu v prostoru i

3. Vizualní odometrie

posun mezi snímky. Ve speciálních případech mohou být používány i systémy pracující s více než dvěma kamerami.

Dále se metody dělí na příznakové a globální podle způsobu, jakým počítají změnu mezi jednotlivými snímky. Příznaky (features) jsou význačné body v obraze, které lze snadno detekovat a přiřadit jim jednoznačný identifikátor (deskriptor). Příznaky reprezentují obrazy skutečných bodů v prostoru a je na jejich základě možné určovat korespondující si body mezi snímky, které odpovídají obrazu stejného bodu v prostoru. Příznakové metody jsou založené na hledání stejných příznaků mezi posunutými snímky, jehož výsledkem je množina dvojic korespondujících si bodů, ze které lze spočítat výsledný posun. Oproti tomu globální metody pracují přímo s intenzitou nebo barvou jednotlivých pixelů, na základě které se snaží určit posun snímků vůči sobě. K přiřazování zpravidla využívají korelaci. Příznakové metody jsou obecně oproti globálním metodám robustnější a méně výpočetně náročné.

Příznakové metody se dále liší způsobem výpočtu posunu mezi po sobě následujícími snímky. Tyto postupy mohou být označovány jako „2D do 2D“, „3D do 3D“ a „3D do 2D“, podle způsobu reprezentace bodu v přechodím a aktuálním snímku, pro který je počítán posun. Výpočet posunu probíhá z korespondencí bodů v přechodím a aktuálním snímku. Body si korespondují, pokud reprezentují stejný bod v prostoru snímaném kamerou. V případě systému pouze s jednou kamerou je bod reprezentován 2D souřadnicemi v obraze kamery. U stereo kamery může být reprezentován i jako 3D bod v prostoru.

U metody „2D do 2D“ jsou body reprezentovány 2D souřadnicemi a pro výpočet posunu je potřeba alespoň 5 dvojic korespondujících bodů mezi snímky. U takto vypočtené translace ale nelze jednoznačně určit měřítko.

Metoda „3D do 3D“ používá výhradně 3D reprezentace bodů a pro určení posunu mezi snímky stačí pouze 3 korespondence.

U metody „3D do 2D“ je 3D bod určen pouze v předchozím snímku. V aktuálním snímku je bod reprezentován 2D obrazem bodu. Posun lze spočítat ze 4 korespondencí mezi 2D a 3D množinami bodů. Jednotlivé příznakové metody jsou podrobněji popsány v kapitole 3.3.

3.2 Geometrie kamery

3.2.1 Model perspektivní kamery

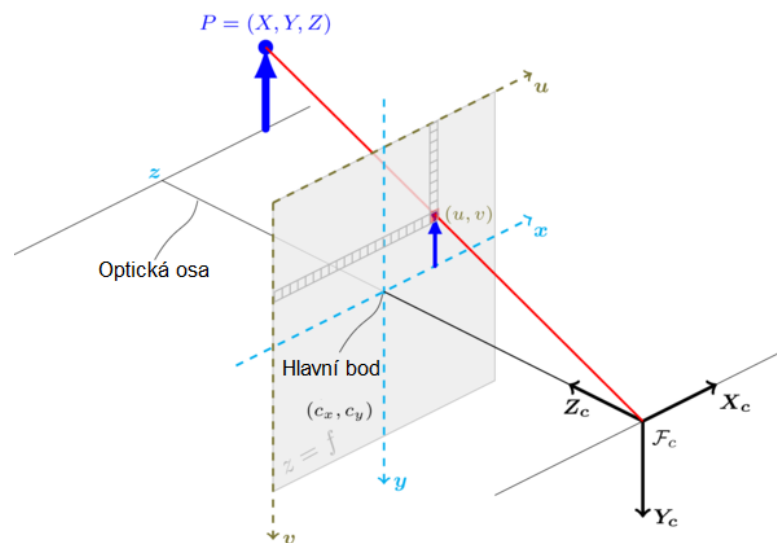
Model perspektivní kamery je matematickým modelem, který umožňuje výpočet souřadnic bodu v obraze při známé pozici zobrazovaného bodu a kamery v prostoru. Někdy může být označován jako model dírkové komory (pinhole camera model). Při výpočtech dochází postupně k transformacím mezi třemi souřadnými systémy (Obr. 3).

První souřadný systém popisuje trojrozměrný prostor snímaný kamerou, který je definován počátečním bodem O a ortonormální bází β . Obvykle se jedná o kartézskou soustavu souřadnic s jednotkami v metrické soustavě. Systém slouží pro popis polohy obecného bodu v prostoru. Souřadnice bodu budou značeny vektorem $\vec{X} = [X, Y, Z]^T$.

Pozice obrazu bodu v promítací rovině je popsána dvourozměrným souřadným systémem, který je určen počátečním bodem a dvěma ortogonálními bázovými vektory. Obecně lze mít i modely, ve kterých tyto vektory na sebe kolmé nebudou, ale v tom případě by nebylo možné využít některé vztahy pro převod souřadnic, které jsou uvedeny dále v této práci. Souřadnice bodu nemusí být udávány v jednotkách používaných pro popis bodu v prostoru, ale mohou být

vyjádřeny i v pixelech. Souřadnice obrazu bodu budou značeny u a v a pozice bodů budou značeny vektorem \vec{u} .

Poslední systém je souřadný systém kamery, který je mezikrokem při převodu mezi předchozími dvěma popisovanými systémy. Opět popisuje trojrozměrný prostor, který je ale pevně svázán s kamerou. Jeho souřadnice budou značeny x , y , z . Počátek je umístěn ve středu promítání F_C a souřadnice z má směr ze středu promítání do hlavního bodu v promítací rovině. Hlavní bod je průsečíkem promítací roviny s přímkou, která je na rovinu promítání kolmá a prochází středem promítání. Vektory souřadnic x a y jsou rovnoběžné se souřadnicemi u a v ze souřadného systému obrazu. Bázové vektory tohoto systému (F_C , β') jsou zároveň rotací vektorů báze β .



Obr. 3 Perspektivní model kamery se znázorněním souřadných systémů, zdroj [4]

Vztah mezi zavedenými souřadnými systémy lze vyjádřit následující rovnicí:

$$\lambda \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.1)$$

Na levé straně této rovnice je vektor pozice obrazu bodu v promítací rovině v homogenních souřadnicích násoben volným parametrem λ . Na pravé straně je pozice bodu v prostoru zapsaná v homogenních souřadnicích násobena zleva maticí pro převod do souřadného systému kamery, která se skládá z rotace a translace. Bod v souřadném systému kamery je dále násoben zleva maticí kamery K . Prvky této matice se nazývají vnitřními parametry kamery a bývají zpravidla neměnné. Pokud by v souřadném systému obrazu byly používány stejné jednotky jako v souřadném systému kamery, parametry f_x a f_y by představovaly vzdálenost středu promítání od hlavního bodu. Obvykle jsou hodnoty obrazových souřadnic udávány v pixelech. Parametry c_x a c_y reprezentují souřadnice hlavního bodu v obrazových souřadnicích. Pokud je známá transformace mezi souřadným systémem kamery a prostorovými souřadnicemi popisujícími pozice bodů a matice kamery, lze spočítat obrazové souřadnice libovolného snímaného bodu. Výsledkem roznásobení pravé strany rovnice je třírozměrný vektor, ze kterého se obrazové souřadnice získají vydělením celého vektoru posledním prvkem, který představuje velikost parametru λ .

3. Vizuální odometrie

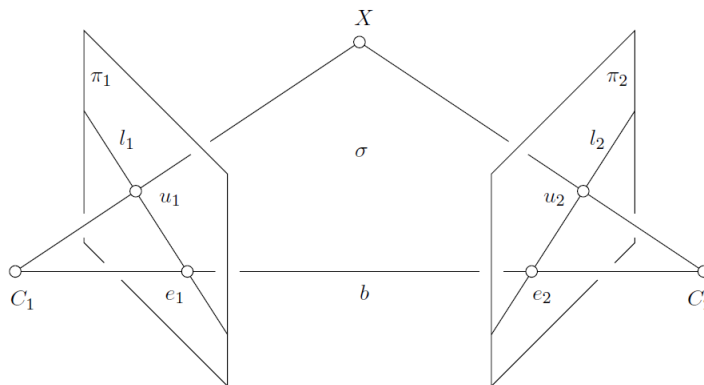
Opačný převod z obrazových souřadnic do prostorových je možný také, ale výsledkem je místo konkrétního bodu v prostoru celá přímka, jejíž body se zobrazují do daného bodu v obraze. Pokud bychom chtěli znát přesnou pozici daného bodu, museli bychom použít snímání více kamerami (viz kapitola 3.2.2) nebo snímat alespoň čtyři body se známou pozicí vůči sobě.

3.2.2 Stereo kamera

3.2.2.1 Epipolární geometrie

Jak již bylo zmíněno v předchozí kapitole, při použití pouze jedné kamery není možné z obrazu bodu jednoznačně určit jeho polohu v prostoru. Proto se často využívá stereo kamera, která je složena ze dvou pevně spojených kamer s různými středy promítání. Geometrie stereo kamery, která sestává ze dvou perspektivních modelů kamery, je znázorněna na Obr. 4. Pro vzájemnou geometrii dvou kamer se používá termín epipolární geometrie.

Významným bodem v promítací rovině každé kamery je epipól, který je obrazem středu druhé kamery. Pro každý bod \vec{u} v promítací rovině kamery lze určit přímku, na které leží všechny body \vec{X} v prostoru, které se zobrazí do daného bodu. Protože obrazem přímky v perspektivním promítání je opět přímka, zobrazí se ve druhé kameře množina všech bodů v prostoru s obrazem v bodě \vec{u} opět na přímce, která je nazývána epipolární přímkou (též epipolára). Této vlastnosti stereo kamery využívají některé metody pro hledání korespondujících bodů v obraze, kdy je podél epipolární přímky hledán bod, který je nejpodobnější zvolenému bodu v obraze druhé kamery. Všechny epipolární přímky prochází společným epipólem. Speciálním případem konfigurace kamer je, pokud je spojnice středů promítání rovnoběžná k promítací rovině některé z kamer. Potom není možné určit epipól, protože spojnice neprotne promítací rovinu. Epipolární přímky však určit možné je a platí pro ně, že jsou rovnoběžné se spojnicí středů promítání.



Obr. 4 Znárodnění epipolární geometrie kamery. Body C_1, C_2 značí středy promítání kamer, e_1 a e_2 jsou epipóly, přímky l_1 a l_2 představují epipolární přímky a body u_1, u_2 znázorňují obrazy bodu X , který je určen prostorovými souřadnicemi v soustavě σ . π_1 a π_2 jsou roviny promítání. (zdroj [5], str. 114)

Pro popis epipolární geometrie se využívá elementární matice E a fundamentální matice F . Jejich odvození je možné najít v [5], str. 113 - 116. Pro jejich popis slouží rotační matice R_1 a R_2 , které jsou rotací prostorových souřadnic σ do souřadnic kamery, středy promítání obou kamer C_1 a C_2 v souřadné soustavě σ a matice kamery K_1 a K_2 . Elementární matice je definována jako:

$$E = R_2 \cdot [C_2 - C_1]_{\times} \cdot R_1^T \quad (3.2)$$

Značení $[\vec{v}]_{\times}$ slouží pro maticový zápis vektorového součinu, který lze spočítat vynásobením této matice zprava druhým z vektorů zapsaným ve sloupcové matici. Ve fundamentální matici jsou navíc obsaženy i matice kamer.

$$F = K_2^{-T} \cdot R_2 \cdot [C_2 - C_1]_{\times} \cdot R_1^T \cdot K_1^{-1} \quad (3.3)$$

Pokud se bod \vec{X} zobrazí do bodů \vec{u}_1 a \vec{u}_2 v homogenních obrazových souřadnicích kamer, platí pro ně vztah:

$$\vec{u}_2^T \cdot F \cdot \vec{u}_1 = 0 \quad (3.4)$$

Dále lze z fundamentální matice přímo určit epipolární přímky, které jsou reprezentovány vektory \vec{l}_1, \vec{l}_2 .

$$\vec{l}_1 = F^T \cdot \vec{u}_2, \quad \vec{l}_2 = F \cdot \vec{u}_1 \quad (3.5)$$

Přímka je určena vektorem tak, že pro body \vec{u} , které na ní leží, musí platit:

$$\vec{l}^T \cdot \vec{u} = 0 \quad (3.6)$$

3.2.2.2 Výpočet pozice bodu ze souřadnic v obraze

Pro určení pozice bodu v 3D prostoru z jeho obrazů je potřeba znát pozice kamer v prostorových souřadnicích, které jsou reprezentovány maticemi složenými z rotační matice a vektoru translace, a matice kamery K_1 a K_2 . Často nejsou pozice kamer určeny v obecných prostorových souřadnicích, ale pouze relativně vůči sobě pomocí matice rotace R a vektoru posunu \vec{t} , které určují transformaci mezi souřadnými systémy kamer. V tomto případě se zpravidla volí matice P_1 a P_2 ve tvaru:

$$P_1 = K_1 \cdot \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \end{bmatrix}, \quad P_2 = K_2 \cdot [R|\vec{t}] \quad (3.7)$$

Rovnici (3.1) lze přepsat do tvaru s maticí P pro přímou transformaci z prostorových do obrazových souřadnic, která bude dále rozdělena na jednotlivé řádkové vektory:

$$\lambda \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \vec{p}_1^T \\ \vec{p}_2^T \\ \vec{p}_3^T \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.8)$$

Pro první kameru budou obrazové souřadnice značeny u, v a řádkové vektory matice P \vec{p}_1^T, \vec{p}_2^T a \vec{p}_3^T . Tyto parametry pro druhou kameru budou značeny přidáním čárkové notace. Úpravami této rovnice pro obě kamery a jejich spojením je možné odvodit matici A , ze které lze spočítat souřadnice bodu \vec{X} v prostoru řešením následující lineární homogenní rovnice.

$$A \cdot \begin{bmatrix} X_{4D} \\ Y_{4D} \\ Z_{4D} \\ W_{4D} \end{bmatrix} = \begin{bmatrix} u \cdot \vec{p}_3^T - \vec{p}_1^T \\ v \cdot \vec{p}_3^T - \vec{p}_2^T \\ u' \cdot \vec{p}'_3 - \vec{p}'_1 \\ v' \cdot \vec{p}'_3 - \vec{p}'_2 \end{bmatrix} \cdot \begin{bmatrix} X_{4D} \\ Y_{4D} \\ Z_{4D} \\ W_{4D} \end{bmatrix} = \vec{0} \quad (3.9)$$

3. Vizualní odometrie

Při výpočtech souřadnic bodu \vec{X} je nejprve vypočítán čtyřrozměrný vektor \vec{X}_{4D} , který představuje homogenní souřadnice bodu \vec{X} vynásobené skalárem. Vydělením tohoto vektoru posledním prvkem, který musí mít v homogenních souřadnicích velikost 1, je možné získat výsledné souřadnice.

$$X = \frac{X_{4D}}{W_{4D}}, \quad Y = \frac{Y_{4D}}{W_{4D}}, \quad Z = \frac{Z_{4D}}{W_{4D}} \quad (3.10)$$

Pozici bodu v prostoru není možné počítat přímo dosazením vektoru \vec{X} v homogenních souřadnicích, protože při použití reálné kamery není možné změřit pozice bodů úplně přesně. Výsledná soustava má pouze 3 neznámé na 4 rovnice a vlivem chyby měření by téměř nikdy neměla řešení. Proto je nezbytné použít některou z metod pro řešení přeuročené soustavy rovnic.

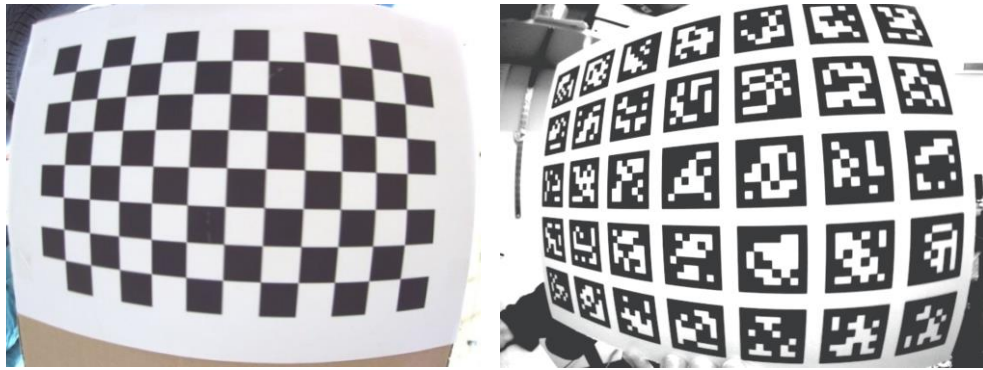
Metoda prezentovaná v této kapitole je používána i knihovnou OpenCV [8]. Její nevýhodou je, že není invariantní vůči transformaci projekce. Výhodou je naopak snadná rozšiřitelnost pro metody, které používají více než dvě kamery.

3.2.3 Kalibrace kamery

V předchozích dvou kapitolách byly popsány vztahy mezi obrazem bodu v obrazové rovině kamery a jeho pozicí v prostoru, ve kterých se však vyskytovalo mnoho parametrů, které musí být známy, aby bylo možné určit pozici bodu v prostoru nebo v obraze kamery. Těmito parametry jsou některé prvky matice kamery, transformační matice do souřadného systému kamery, vzájemná pozice kamer ve stereo páru nebo fundamentální a elementární matice. Jednou z možností, jak tyto parametry určovat, je dosadit hodnoty do matic přímo na základě znalosti rozmístění kamer a jejich vlastností. Při reálných měřeních ale obvykle není možné tyto parametry určit přímo s dostatečnou přesností, a proto jsou obvykle hledány metodami pro kalibraci kamer, které dokáží určit tyto parametry z obrazové informace kalibrovaných kamer. Kalibrace probíhá obvykle snímáním speciální značky se známými rozměry, ze kterých je možné dopočítat jednotlivé parametry.

Metody pro kalibraci kamer jsou obsaženy i v knihovně OpenCV [4]. Pro kalibraci se používá rovinná značka se vzorem šachovnice. Výstupem kalibrace jedné kamery jsou kromě samotné matice kamery i distorzní koeficienty, které reprezentují optické vady kamery. Tyto koeficienty umožňují odstranit zkreslení snímků a následně použití perspektivního modelu kamery. Počet distorzních koeficientů může být 2-6 pro radiální zkreslení a 2 pro tangenciální zkreslení. Při kalibraci stereo páru je výstupem rotační matice R a vektor posunu \vec{t} pro transformaci mezi souřadným systémem první a druhé kamery. Dále funkce vrací fundamentální i esenciální matici (viz kapitola 3.2.2.1).

Dalším z nástrojů pro kalibraci kamery je APRIL camera calibration suite [9]. Značka pro kalibraci sestává z matice jednotlivých AprilTagů. AprilTagy [10] jsou čtvercové značky, které mohou být snadno detekovány v obraze z kamery. Z detekované značky lze určit její polohu, orientaci a identifikátor, který umožňuje rozlišit jednotlivé značky. Oproti kalibraci v OpenCV je možné provádět kalibraci i ze snímků, ve kterých není zachycena celá matice značek. Tato vlastnost velmi usnadňuje kalibraci kamery a snižuje riziko pořízení chybných dat. Dále je APRIL kalibrace vybavena nástrojem *AprilCal*, který prostřednictvím grafického rozhraní provází uživatele kalibrací a doporučuje, ve které pozici značky by měl být pořízen další snímek pro přesnější kalibraci. Stereo pár je možné kalibrovat nástrojem *MultiCameraCalibrator*.



Obr. 5 Snímky značek pro kalibraci kamery pomocí nástrojů v knihovně OpenCV (vlevo) a APRIL kalibrace (vpravo, zdroj [9])

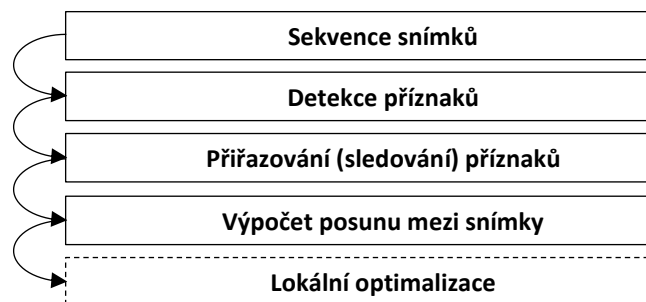
3.3 Analýza postupů s využitím příznaků

3.3.1 Výpočetní vlákno příznakové vizuální odometrie

Jak bylo již zmíněno v kapitole 3.1, cílem vizuální odometrie je určit trajektorii kamery na základě změny v obraze z kamery. Vstupem systému jsou jednotlivé snímky, které budou značeny I_i , kde index i je nezáporné celé číslo udávající pořadí snímku. Snímky vstupují do systému postupně a vždy je pro aktuální snímek určen posun vůči předchozím snímkům. Zpravidla je počítán pouze relativní posun mezi aktuálním a předchozím snímkem, který je vyjádřen maticí $T_{i,i-1}$ sestávající z rotační matice a vektoru translace. Výsledná pozice vůči počátečnímu snímku je získána spojením aktuálního posunu se všemi předchozími.

$$T_{i,i-1} = \begin{bmatrix} R_{i,i-1} & \vec{t}_{i,i-1} \\ 0 & 1 \end{bmatrix} \quad (3.11)$$

Výpočetní vlákno vizuální odometrie se obecně skládá z několika částí, které jsou znázorněné v blokovém diagramu na Obr. 6.



Obr. 6 Blokový diagram vlákna pro vizuální odometrii pracující s příznaky (zdroj [2])

Jeho vstupem je sekvence po sobě následujících snímků nebo dvojic snímků v případě stereo kamery. Další částí je detekce příznaků v jednotlivých snímcích. Po detekci následuje přiřazování nebo sledování příznaků. Při přiřazování jsou příznaky hledány v jednotlivých snímcích nezávisle na sobě. Dále jsou mezi po sobě těmito množinami příznaků hledány korespondující si dvojice bodů. Oproti tomu při sledování je odhadnuta pozice, na které by se měl vyskytovat korespondující příznak a v tomto místě je hledán pomocí nástrojů pro lokální prohledávání. Na základě korespondujících si dvojic bodů je počítán posun mezi následujícími snímky. Jednotlivé metody výpočtu jsou podrobněji rozebrány v navazujících kapitolách. Poslední částí, kterou nemusí obsahovat všechny systémy, je lokální optimalizace, při které dochází ke zpřesnění výsledné pozice na základě dat z více předchozích snímků.

3.3.2 2D do 2D

Při výpočtu posunu metodami 2D do 2D systém zpravidla využívá pouze jednu pohybující se kameru. Posun mezi snímky I_i a I_{i-1} , kde i značí pořadí snímku, je určován výpočtem esenciální matice, ze které je možné určit rotaci i translaci. Vektor translace ale nemá jednoznačně určené měřítko. Samotnému výpočtu předchází přiřazení příznaků, které odpovídají stejným bodům v prostoru mezi oběma snímky. Výstupem přiřazení je množina korespondujících si dvojic. V této kapitole je dále předpokládáno, že je tato množina známá a neobsahuje chybná přiřazení.

3.3.2.1 Osmibodová metoda

Pro esenciální matici E platí, že:

$$E = \lambda \cdot [\vec{t}]_x \cdot R \quad (3.12)$$

kde λ je reálné číslo, které reprezentuje nejednoznačnost měřítka. Rotační matice R a vektor translace \vec{t} představují posun snímku I_i vůči snímku I_{i-1} . Základní metodou pro výpočet esenciální matice je osmibodová metoda (článek [2], str. 87), která vyžaduje alespoň osm korespondencí mezi snímky.

Body ve snímku I_{i-1} budou značeny $\vec{p}_1 = [u_1 \ v_1 \ 1]^T$ a pro korespondující body ve snímku I_i bude používáno podobné značení $\vec{p}_2 = [u_2 \ v_2 \ 1]^T$. Nalezené body musí splňovat následující epipolární podmínku, pro kterou lze matici E rozdělit na jednotlivé prvky.

$$p_2^T \cdot E \cdot p_1 = 0 \rightarrow [u_2 \ v_2 \ 1] \cdot \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0 \quad (3.13)$$

Tuto podmínku lze upravit tak, aby matice E byla sloupcový vektor ve tvaru $E_C = [e_1 \ \dots \ e_9]^T$:

$$[u_1 u_2 \ v_1 u_2 \ u_2 \ u_1 v_2 \ v_1 v_2 \ v_2 \ u_1 \ v_1 \ 1] \cdot E_C = 0 \quad (3.14)$$

Z korespondence 8 bodů může být vytvořena matice A s rozměry 8×9 , která vznikne spojením levých matic z rovnice (3.14) pro jednotlivé korespondence. Výsledkem je homogenní soustava lineárních rovnic ve tvaru:

$$A \cdot E_C = \vec{0} \quad (3.15)$$

Pro řešení této rovnice se využívá metoda singulárního rozkladu matice (SVD) pro matici A , který vrátí matice U , S a V .

$$A = U \cdot S \cdot V^T \quad (3.16)$$

Matici E_C lze získat z posledního sloupce matice V . Tento odhad metodou nejmenších čtverců po převodu zpět na čtvercovou matici nesplňuje některé podmínky pro esenciální matici. Pro vlastní čísla esenciální matice platí, že má dvě vlastní čísla shodná a třetí je rovné nule. Tato vlastnost bude zajištěna tak, že po aplikaci SVD na matici E_C , dojde k výměně matice S za diagonální matici s vlastními čísly 0, 1, 1. Tím je získána výsledná esenciální matice E .

$$E_C = U \cdot S \cdot V^T \rightarrow E = U \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot V^T \quad (3.17)$$

Výhodou osmibodové metody je, že může pracovat i s kamerou, která není kalibrována. Naopak tato metoda selhává, pokud 3D body v prostoru reprezentované korespondujícími si body leží v jedné rovině.

3.3.2.2 Pětibodová metoda

Minimální počet korespondencí pro získání esenciální matice je pět, čehož využívá pětibodová metoda [12]. Tato metoda funguje i pro obrazy bodů ležících v jedné rovině, ale vyžaduje naopak kalibraci kamery. Body v souřadném systému kamery pro jednotlivé pohledy budou značeny \vec{q}_1, \vec{q}_2 . Převod z obrazových souřadnic do souřadnic kamery je prováděn pomocí inverzní matice k matici kamery K . Obdobně, jako u osmibodové metody, lze přepsat podmínku pro esenciální matici pomocí sloupcové matice E_C .

$$[q_{1x}q_{2x} \ q_{1y}q_{2x} \ q_{1z}q_{2x} \ q_{1x}q_{2y} \ q_{1y}q_{2y} \ q_{1z}q_{2y} \ q_{1x}q_{2z} \ q_{1y}q_{2z} \ q_{1z}q_{2z}] \cdot E_C = 0 \quad (3.18)$$

Z korespondencí pěti bodů se dále vytvoří matice A sjednocením matic na levé straně předchozí rovnice. Řešením takto vzniklé soustavy rovnic $A \cdot E_C = 0$ je čtyřrozměrný prostor určený vektory $\vec{x}, \vec{y}, \vec{z}, \vec{w}$, které lze získat pomocí SVD nebo QR-faktorizací. Tyto vektory o devíti prvcích lze převést na matice o rozměru 3×3 . Jde o inverzní operaci k vytvoření sloupcové matice z matice E . Následně lze matici E zapsat pomocí nově získaných matic X, Y, Z, W a reálných čísel x, y, z, w .

$$E = x \cdot X + y \cdot Y + z \cdot Z + w \cdot W \quad (3.19)$$

Protože je měřítko matice E nejednoznačné, může být zvolena hodnota parametru $w=1$. Pro esenciální matici platí následující dvě omezení, která jsou aplikována na vztah (3.19).

$$\begin{aligned} \det(E) &= 0 \\ EE^T E - \frac{1}{2} \text{trace}(EE^T)E &= 0 \end{aligned} \quad (3.20)$$

Z těchto podmínek vychází 10 omezení na hodnoty x, y a z , které jsou reprezentovány soustavou 10 kubických rovnic, ve kterých je jedinou neznámou z . Tyto rovnice lze dále upravit na soustavu reprezentovanou maticí B , která obsahuje polynomy proměnné z až 4. řádu. Aby měla tato soustava řešení, musí platit, že determinant matice B je roven 0. Determinant představuje polynom 10. řádu, pro který lze hledat kořeny pomocí Strumovy posloupnosti nebo sdružené matice. Z nalezených řešení pro hodnotu z lze dosazením do matice B spočítat hodnoty x a y . Výsledná matice E se získá dosazením těchto reálných čísel do rovnice (3.19). Podrobnější popis této metody lze nalézt v článku [12].

3.3.2.3 Výpočet rotace a translace z esenciální matice

Pro získání matice rotace a translace z matice E se využívá jejího rozkladu pomocí SVD, který vrací matice U, S a V . Dále budou zavedeny 4 matice W .

$$\begin{aligned} W_1 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & W_2 &= \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ W_3 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, & W_4 &= \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \end{aligned} \quad (3.21)$$

3. Vizualní odometrie

Nejprve jsou získána 4 různá řešení pro R a \vec{t} pomocí matic W .

$$\begin{aligned} R_x &= UW_x^T V^T \\ \vec{t}_x &= UW_x S U^T \end{aligned} \quad (3.22)$$

Správné řešení pro R a t lze získat triangulací jednoho korespondujícího bodu pomocí všech 4 nalezených řešení, kde pouze pro správné řešení bude platit, že se triangulovaný bod nachází před oběma kamerami.

Tímto postupem lze získat transformace mezi dvojicemi po sobě jdoucích snímků. Nicméně pro získání výsledné polohy vůči počáteční poloze nelze takto nalezené transformace přímo spojovat, protože mohou mít jiné měřítko, které není určeno jednoznačně. Základní metodou pro sjednocení měřítka je výpočet parametru r . Tento parametr převádí nově získanou transformaci na předchozí porovnáním vzdáleností pozic dvou stejných bodů \vec{X}_1, \vec{X}_2 triangulovaných v předchozím a novém snímku.

$$r = \frac{\|\vec{X}_{2,i-1} - \vec{X}_{1,i-1}\|}{\|\vec{X}_{2,i} - \vec{X}_{1,i}\|} \quad (3.23)$$

Pro přesnější výsledek je možné spočítat parametr r pro více vzdáleností a zprůměrovat je. Výsledný vektor translace je vynásobením původně vypočteného vektoru parametrem r .

3.3.3 3D do 3D

Vstupem metody 3D do 3D je posloupnost dvojic snímků ze stereo kamery, ze kterých lze triangulací získat množinu 3D bodů v prostoru. Tyto body jsou navíc reprezentovány deskriptorem příznaku, a proto mohou být přiřazovány mezi jednotlivými kroky. Pro další popis metody bude předpokládáno, že body z po sobě následujících stereo snímků byly triangulovány, přiřazeny a z přiřazených hodnot byly odfiltrovány chybné hodnoty. Výstupem těchto kroků je množina bodů $\{\vec{X}_1, \dots, \vec{X}_2\}$, pro které jsou známy jejich souřadnice v souřadných systémech kamery pro pozice i a $i-1$. Úloha hledání matice R a vektoru \vec{t} reprezentujících pohyb mezi snímky může být formulována jako hledání optimálního řešení pomocí metody nejmenších čtverců. Pro homogenní souřadnice bodu \vec{x} v souřadném systému kamery a matici posunu T , která je spojením rotace a translace lze úlohu definovat následujícím způsobem:

$$\operatorname{argmin}_T \sum_k \|\vec{x}_{k,i} - T \cdot \vec{x}_{k,i-1}\|, \quad k = 1, \dots, N \quad (3.24)$$

Pro nalezení řešení jsou potřeba alespoň 3 korespondence, pro které musí platit, že tyto body v prostoru neleží v jedné přímce.

Jedním z možných řešení této úlohy je použití výpočtu těžiště množiny bodů [13], jehož souřadnice jsou průměrnými souřadnicemi všech bodů množiny. Pro body ze snímků v kroku i a $i-1$ budou body těžiště značeny \bar{x}_i a \bar{x}_{i-1} .

Jako první bude určována rotační matice R . Z množin korespondujících si bodů v obou snímcích a jejich těžišť se určí matice H , pro kterou se provede singulární rozklad na matice U, V a S .

$$H = \sum_k (\vec{x}_{k,i-1} - \bar{x}_{i-1}) \cdot (\vec{x}_{k,i} - \bar{x}_i)^T = USV^T, \quad k = 1, \dots, N \quad (3.25)$$

Matici R lze nyní získat z vypočtených hodnot.

$$R = VU^T \quad (3.26)$$

Matice R by měla splňovat podmínku, že determinant R je roven jedné (s tolerancí chyby, která je způsobena chybou měření a numerickými chybami při výpočtu). Pokud by vyšel determinant -1 , není výsledek validní. Vektor posunu \vec{t} může být určen přímo z rotační matice R a bodů reprezentujících těžiště.

$$\vec{t} = \bar{x}_i - R \cdot \bar{x}_{i-1} \quad (3.27)$$

Protože metoda 3D do 3D určuje posun v 6DoF a není zde nejistota v měřítku, mohou být vypočtené transformace mezi jednotlivými snímky přímo skládány pro výpočet absolutní pozice ve stanovené souřadné soustavě.

3.3.4 3D do 2D

Přestože vstupem metody 3D do 2D zpravidla bývá posloupnost dvojic snímků ze stereo kamery, využívá se triangulace pouze u snímků v kroku $i-1$. K 3D bodům získaným z triangulace jsou následně přiřazovány body odpovídající jejich průmětům v jednom ze snímků v čase i . Posun mezi snímky je hledán tak, aby byla co nejmenší odchylka mezi teoreticky určenými průměty \vec{p}_{i-1} 3D bodů \vec{X}_{i-1} transformovaných o nalezený posun T a naměřenými průměty \vec{p}_i . Tuto úlohu lze formulovat následujícím způsobem:

$$\operatorname{argmin}_T \sum_k \|\vec{p}_{k,i} - \vec{p}_{k,i-1}\|^2, \quad k = 1, \dots, N \quad (3.28)$$

Tato úloha bývá označována jako problém n bodů, který je v zahraniční literatuře označován zkratkou PnP (perspective of n points). Problém tří bodů, který využívá nejmenšího možného počtu bodů, které navíc nesmí ležet v jedné přímce, má 4 různá řešení. Správné řešení lze vybrat na základě průmětu dalšího bodu. Pro jednoznačné určení polohy je proto zapotřebí alespoň 4 bodů. Při použití 6 a více korespondencí je možné řešit tuto úlohu jako soustavu lineárních rovnic. Oba zmíněné přístupy budou popsány v následujících kapitolách.

3.3.4.1 P6P

V úloze problému 6 bodů ([6], str. 178) je hledána matice P , která je spojením rotační matice a vektoru translace, ve tvaru $P = [R \ \vec{t}]$. Pro 3D body ze snímků v čase $i-1$ a jejich obrazy v čase i bude platit následující vztah.

$$\lambda K^{-1} \vec{u} = \lambda K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P \cdot \vec{X} \quad (3.29)$$

Vektor \vec{X} reprezentuje souřadnice 3D bodu a \vec{u} jeho obraz v obrazových souřadnicích. V dalších výpočtech bude $K^{-1} \vec{u}$ nahrazeno vektorem \vec{x} . Při odvození rovnic pro problém 6 bodů je využito vlastnosti, že vektor \vec{x} odpovídá $P \cdot \vec{X}$ až na měřítko reprezentované parametrem λ a jejich vektorový součin proto bude roven nule. Z upravené rovnice pro vektorový součin, ve které bude matice P reprezentována sloupcovou maticí P_c s rozměry 12×1 , lze získat výslednou rovnici pro řešení problému 6 bodů. Prvky vektoru \vec{x} budou značeny x, y, z .

3. Vizualní odometrie

$$\begin{bmatrix} \vec{o}^T & -z_i \vec{X}_i^T & y_i \vec{X}_i^T \\ z_i \vec{X}_i^T & \vec{o}^T & -x_i \vec{X}_i^T \end{bmatrix} \cdot P_C = \vec{o} \quad (3.30)$$

Spojením těchto matic pro 6 různých bodů \vec{X} a jejich průmětů vznikne výsledná matice pro rovnici ve tvaru $AP_C = 0$. Soustavu lze řešit například metodou SVD.

Pro získání přesnějších výsledků se mohou hodnoty vektorů \vec{x} a \vec{X} normalizovat. Normalizaci bodů \vec{x} , které leží v jedné rovině, je doporučeno provádět tak, aby těžiště bodů bylo v počátku a průměrná vzdálenost bodů od počátku byla $\sqrt{2}$. Pro 3D pozice bodů \vec{X} je způsob normalizace vybírán podle jejich rozdělení v prostoru. Pokud jsou body relativně blízko sebe, doporučuje se podobný přístup jako pro 2D souřadnice, tedy umístění těžiště do středu a aby průměrná souřadnice byla $[1, 1, 1, 1]^T$. Obě tyto normalizace jsou reprezentovány maticemi N_V a N_U .

$$\vec{x}_n = N_V \cdot \vec{x}, \quad \vec{X} = N_U \cdot \vec{X} \quad (3.31)$$

Výslednou matici P lze z matice nalezené pro normalizované hodnoty získat vynásobením inverzní maticí k N_V a maticí N_U .

$$P = N_V^{-1} \cdot P_n \cdot N_U \quad (3.32)$$

3.3.4.2 P3P

Jedno z možných řešení problému tří bodů je popsáno v článku [14]. Vstupem jsou tři průměty bodů v obrazových souřadnicích kamery a jejich 3D pozice vůči obecným prostorovým souřadnicím. Výstupem bude pozice kamery v prostorových souřadnicích, která bude popsána rotační maticí R a translací \vec{t} . Protože jsou použity pouze tři body, existují zpravidla čtyři různá řešení. Pro výběr správného řešení je zapotřebí znalost 3D pozice a průmětu čtvrtého bodu. Při použití metody pro úlohu vizuální odometrie, budou prostorovými souřadnicemi souřadnice stereo kamery pro snímek v čase $i-1$.

V prvním kroku je vypočtena transformace souřadného systému kamery v do nového souřadného systému τ . Pro výpočet jsou používány jednotkové vektory $\vec{f}_1, \vec{f}_2, \vec{f}_3$ v souřadnicích kamery se směrem ze středu kamery C do průmětů bodů. Směry těchto vektorů odpovídají směrům, ve kterých se nacházely skutečné body v 3D prostoru. Vektory nového souřadného systému se následně vypočítají z vektorů \vec{f} .

$$\vec{t}_x = \vec{f}_1, \quad \vec{t}_z = \frac{\vec{f}_1 \times \vec{f}_2}{\|\vec{f}_1 \times \vec{f}_2\|}, \quad \vec{t}_y = \vec{t}_z \times \vec{t}_x \quad (3.33)$$

Spojením vektorů do matice $T = [\vec{t}_x \quad \vec{t}_y \quad \vec{t}_z]^T$ vznikne matice transformující body \vec{f}_i ze souřadného systému kamery v do τ .

$$\vec{f}_i^\tau = T \cdot \vec{f}_i \quad (3.34)$$

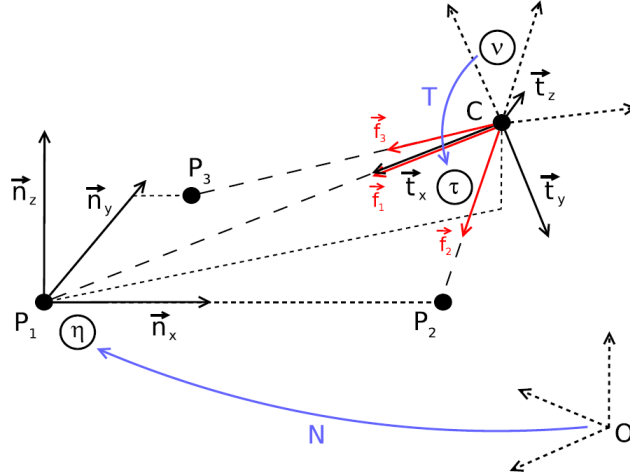
Obdobně je vytvořen nový souřadný systém η pro prostorové souřadnice, který je definován pomocí 3D bodů $\vec{P}_1, \vec{P}_2, \vec{P}_3$.

$$\vec{n}_x = \frac{\overrightarrow{P_1 P_2}}{\|\overrightarrow{P_1 P_2}\|}, \quad \vec{n}_z = \frac{\vec{n}_x \times \overrightarrow{P_1 P_3}}{\|\vec{n}_x \times \overrightarrow{P_1 P_3}\|}, \quad \vec{n}_y = \vec{n}_z \times \vec{n}_x \quad (3.35)$$

Body ze zadaných prostorových souřadnic do souřadného systému η se transformují následujícím vztahem, kde $N = [\vec{n}_x \ \vec{n}_y \ \vec{n}_z]^T$:

$$P_i^\eta = N \cdot (P_i - P_1) \quad (3.36)$$

Dále bude hledána pouze transformace mezi systémy η a τ , ze které lze pomocí předchozích vztahů dopočítat výslednou transformaci mezi souřadným systémem kamery a prostorovými souřadnicemi. Všechny uvedené souřadné systémy jsou znázorněny na Obr. 7.



Obr. 7 Znárodnění souřadných systémů pro řešení problému tří bodů (zdroj [14])

V další části řešení se úloha redukuje na hledání kořenů polynomu čtvrtého stupně, ve kterém je neznámou $\cos(\theta)$. Úhel θ odpovídá úhlu, který svírá rovina určená rovnicí $\vec{n}_z = 0$ s rovinou určenou body P_1, P_2 a C .

$$a_4 \cos(\theta)^4 + a_3 \cos(\theta)^3 + a_2 \cos(\theta)^2 + a_1 \cos(\theta) + a_0 = 0 \quad (3.37)$$

Podrobnější popis celé metody i s odvozením a vyjádřením jednotlivých parametrů a je možné dohledat v článku [14]. Ze získaných hodnot pro $\cos(\theta)$ lze spočítat $\cotg(\alpha)$, kde úhel $\alpha = \sphericalangle P_2 P_1 C$. Hodnoty dalších goniometrických funkcí úhlů θ a α je možné určit z vypočtených hodnot a známých parametrů. Pozice středu kamery v souřadné soustavě η , která odpovídá translaci při převodu souřadnic, a rotace Q mezi systémy η a τ jsou přímo určeny vypočtenými hodnotami θ a α . Protože polynom 4. řádu může mít čtyři různé kořeny, jsou výsledkem tohoto algoritmu také až čtyři různé transformace.

$$\vec{C}^\eta = \begin{bmatrix} d_{12} \cos(\alpha) (\sin(\alpha) \cdot b + \cos(\alpha)) \\ d_{12} \sin(\alpha) \cos(\theta) (\sin(\alpha) \cdot b + \cos(\alpha)) \\ d_{12} \sin(\alpha) \cos(\theta) (\sin(\alpha) \cdot b + \cos(\alpha)) \end{bmatrix} \quad (3.38)$$

$$Q = \begin{bmatrix} -\cos(\alpha) & -\sin(\alpha) \cos(\theta) & -\sin(\alpha) \sin(\theta) \\ \sin(\alpha) & -\cos(\alpha) \cos(\theta) & -\cos(\alpha) \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

Parametry b a d_{12} jsou určeny přímo vstupními hodnotami úlohy. Parametr d_{12} odpovídá vzdálenosti bodů P_1 a P_2 , b je odvozené z hodnot vektorů \vec{f}_1 a \vec{f}_2 .

$$b = \text{sign}(\vec{f}_1 \cdot \vec{f}_2) \cdot \sqrt{\frac{1}{1 - (\vec{f}_1 \cdot \vec{f}_2)^2} - 1}, \quad d_{12} = \|P_2 - P_1\|, \quad (3.39)$$

3. Vizualní odometrie

Výsledná pozice souřadnic kamery ve snímku i se určí z předchozích hodnot rovnicemi:

$$\vec{t} = P_1 + N^T \cdot C^\eta, \quad R = N^T \cdot Q^T \cdot T \quad (3.40)$$

Na závěr je ze všech nalezených řešení vybráno správné řešení porovnáním zpětné projekce 4. bodu s jeho skutečným obrazem. Výhodou této metody je dobrá numerická stabilita a rychlost výpočtů v porovnání s ostatními metodami pro řešení problému tří bodů. Často se proto využívá v RANSAC algoritmu (random sample consensus). Dále je k dispozici přímo od autorů [14] knihovna s implementovaným algoritmem v jazyce C++.

3.4 Detektory a deskriptory příznaků

V počítačovém vidění existuje několik přístupů k definici a dělení příznaků. V této práci jsou za příznaky považovány pouze lokální fotometrické příznaky podle dělení [16]. Tyto příznaky jsou význačnými body v obraze z kamery, které lze detekovat na základě rozdílné intenzity, barvy nebo textury vůči okolí. Pro každý příznak bývá dále spočítán deskriptor, který slouží jako jednoznačný identifikátor příznaku. Na základě porovnání deskriptorů v různých snímcích lze detekovat příznaky, které odpovídají stejnému bodu v prostoru. Korespondující příznaky je možné najít i ve snímcích pořízených z jiných pohledů nebo při odlišných světelných podmínkách. Čím je ale rozdíl mezi snímky výraznější, tím více klesá pravděpodobnost, že se stejné příznaky podaří detekovat. Pro detekci i výpočet deskriptorů existuje mnoho různých metod, které se liší robustností, počtem i charakterem nalezených příznaků. Každá metoda navíc zpravidla má vlastní vnitřní parametry, které rovněž ovlivňují kvalitu a počet výsledných příznaků.

Dále prezentované metody jsou založené na detekci rohů (corners) nebo kontrastních objektů (blobs). V některých případech se místo rohů detekují hrany, u kterých však bývá obtížné od sebe rozlišit jednotlivé body ležící na stejné hraně. Roh je definován jako průsečík dvou hran a jeho pozice bývá jednoznačná. Kontrastní objekty se liší od okolí intenzitou, barvou nebo texturou. Obecně platí, že výpočet rohů bývá oproti kontrastním objektům rychlejší, ale jednotlivé rohy jsou hůře rozlišitelné. Rohy jsou dále hůře rozpoznatelné při změně rozměrů v obraze vlivem oddálení nebo přiblížení kamery. Naopak jsou ale invariantní vůči rotaci příznaku. V následujících kapitolách budou popsány některé metody pro detekci příznaků a vytváření deskriptorů. Všechny popisované metody jsou implementovány v knihovně OpenCV [8], která je využívána pro detekci a popis příznaků v této práci.

3.4.1 Harrisův detektor

Základním detektorem rohů je Harrisův detektor, který byl představen v článku [19] z roku 1988. Tento algoritmus vychází z prvního detektoru rohů, kterým je Moravcův detektor. Pro porovnávání detekovaných rohů je nezbytné použít některý z deskriptorů příznaků. Moravcův detektor je založen na porovnávání intenzit ve vybraném okně v obraze s jeho posunem určitým směrem, při kterém mohou nastat tři odlišné situace. Ve vybraném okně se může nacházet roh, hrana anebo je oblast prázdná, tedy neobsahuje výrazné změny v intenzitě. Pro oblast bez hran a rohů při posunu okna nedochází k viditelným změnám intenzity. Pro hranu je změna intenzity vůči posunu výrazná, ale pouze ve směru kolmém na hranu. Změna pro posun podél hrany nebývá příliš patrná. Roh je proto možné rozpoznat

jako oblast s výraznou změnou intenzity ve všech směrech. Matematický zápis této myšlenky pro Moravcův operátor je:

$$E_{x,y} = \sum_{u,v} w(u,v) \cdot (I(x+u, y+v) - I(u,v))^2 \quad (3.41)$$

$E_{x,y}$ představuje vypočtený rozdíl pro posun okna ve směru x, y . V Moravcově detektoru jsou předpokládány pouze čtyři posuny ve směrech (1,0), (1,1), (0,1) a (-1,1). Funkce w představuje váhovací funkci pro jednotlivé body okna u, v . Funkce I udává intenzitu v jednotlivých pixelech se zadanými souřadnicemi. V Moravcově detektoru se pro každé okno vybere minimální hodnota funkce E ze všech posunů. Roh je detekován jako oblast, pro kterou nalezené minimum překročí pevně určený práh. Váhovací funkce w ve zvoleném obdélníkovém okně nabývá hodnoty jedna a ve všech ostatních oblastech snímku 0.

Harrisův detektor vylepšuje Moravcův detektor a odstraňuje jeho hlavní nedostatky, kterými jsou anizotropní chování vlivem nízkého počtu diskrétních směrů posunu, šum způsobený binárním obdélníkovým oknem a častá detekce hran zapříčiněná výběrem pouze minimální hodnoty E ze všech rozdílů.

U Harrisova detektoru může být použita i jiná váhovací funkce. Často se používá kruhové okno s váhovací funkcí, která odpovídá normálnímu rozdělení s rozptylem σ :

$$w_{u,v} = e^{-\frac{(u^2+v^2)}{2\sigma}} \quad (3.42)$$

Pro malé posuny okna lze nahradit rozdíl v rovnici (3.41) aproximací parciálních derivací a vztah pro $E_{x,y}$ upravit do tvaru:

$$E_{x,y} = \sum_{u,v} w(u,v) \cdot (X^2x^2 - 2XYxy + Y^2y^2) \quad (3.43)$$

kde $X \approx \frac{\partial I}{\partial x}$ a $Y \approx \frac{\partial I}{\partial y}$.

Tuto rovnici lze dále přepsat do tvaru s maticovým zápisem, ve kterém bude zavedena matice M :

$$E_{x,y} = \sum_{u,v} w(u,v) \cdot [x \ y] \cdot \begin{bmatrix} X^2 & XY \\ XY & Y^2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = [x \ y] \cdot M \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3.44)$$

$$M = \sum_{u,v} w(u,v) \cdot \begin{bmatrix} X^2 & XY \\ XY & Y^2 \end{bmatrix}$$

Matice M může být také označována jako lokální autokorelační funkce a na základě jejích vlastních čísel lze určit, jestli je testovaná oblast rohem, hranou anebo prázdná. Prázdné oblasti mají obě vlastní čísla nízká, zatímco rohy naopak vyšší. Hranu lze poznat podle velkého rozdílu mezi vlastními čísly. Tuto vlastnost zachycuje parametr R , pro který platí:

$$R = \det(M) - k \cdot \text{trace}(M)^2 = \lambda_1\lambda_2 - k \cdot (\lambda_1 + \lambda_2)^2 \quad (3.45)$$

Na základě parametru R lze provést klasifikaci charakteru testované oblasti. Pro oblast, která není rohem ani hranou platí, že absolutní hodnota R je menší, než stanovená hranice. Pro hranu platí, že $R < 0$ a pro rohy naopak $R > 0$. k je volitelný parametr Harrisova detektoru. Obecně nemusí detektor detekovat jen rohy, ale detekuje i oblasti s velkým gradientem.

3.4.2 Shi-Tomasi detektor

Shi-Tomasi detektor je modifikací Harrisova detektoru popsaného v předchozí kapitole, od kterého se odlišuje předpisem pro parametr R . Tento parametr je určen z vlastních čísel matice M . Pokud je hodnota R větší než zadaná hranice, je oblast považována za roh.

$$R = \min(\lambda_1, \lambda_2) \quad (3.46)$$

3.4.3 SIFT

Metoda SIFT (Scale-Invariant Feature Transform) [20] obsahuje detektor i deskriptor kontrastních objektů. Její hlavní předností je detekce příznaku téměř nezávisle na jeho rozměrech, čímž umožňuje přiřazovat příznaky mezi snímky i při velkých pohybech kamery. Tuto metodu lze rozdělit do čtyř částí:

1. Detekce pozic extrémů v různě škálovaném snímku, které jsou určovány diferencí Gaussovy funkce.
2. Lokalizace příznaků na základě extrémů nalezených v předchozím bodě.
3. Přiřazení orientace k příznaku pro zajištění invariance vůči rotaci.
4. Vytvoření deskriptoru příznaku.

Gaussova funkce pro vzdálenosti x, y je určena rovnicí s rozptylem σ :

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma} \cdot e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3.47)$$

Konvolucí Gaussovy funkce pro jednotlivé body obrazu I lze určit funkci L :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.48)$$

Z hlediska změny obrazu se jedná o rozostření pomocí Gaussovy funkce, kde jsou pro daný bod sečteny příspěvky jednotlivých bodů vynásobené Gaussovou funkcí pro vzdálenost přispívajícího bodu od počítaného.

Požadovaná diference Gaussovy funkce lze spočítat jako rozdíl funkcí L pro různé hodnoty rozptylu σ . Funkce L reprezentuje snímek filtrovaný Gaussovou funkcí se zadanou hodnotou σ .

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.49)$$

Této vlastnosti se využívá při samotném výpočtu, kdy jsou jednotlivé filtrované snímky rozděleny na oktávy, pro které platí, že následující oktáva má dvojnásobný rozptyl σ . Pro jednotlivé filtrované snímky platí $k = 2^{1/s}$, kde s udává počet intervalů, na které je oktáva rozdělena. Při rozdělení na s intervalů je pro výpočet extrémů v jedné oktávě zapotřebí $s+3$ snímků. Diference Gaussovy funkce se získá z rozdílu po sobě následujících snímků oktávy. Při přechodu do vyšších oktáv není potřeba počítat jednotlivé snímky znovu, ale lze použít převzorkování snímků z předchozí oktávy tak, že je vybrán pouze každý druhý pixel v každém řádku i sloupci.

Hledání extrémů probíhá v nově získaných snímcích reprezentujících diference v jednotlivých oktávách. Bod je považován za extrém, pokud je extrémem vůči okolním osmi pixelům v daném snímku a 18 pixelům, které se nachází v sousedních snímcích (řazených podle použitého parametru k) na stejné pozici jako daný bod a jeho sousední body.

V dalším kroku jsou nalezené extrémy přesně lokalizovány v obraze a zároveň jsou vyjmuty příliš slabé příznaky nebo příznaky ležící na hranách. Diference v okolí příznaku je aproximována pomocí Taylorova rozvoje 2. řádu a je pro ni nalezen extrém. Pokud je poloha tohoto extrému větší než 0,5, náleží extrém jinému bodu než určenému. Pokud je hodnota diference v nalezeném extrému nižší než zadaná hodnota, je příznak v daném extrému zamítnut jako příliš slabý.

Filtrace hran probíhá na základě určení Hessovy matice podobným způsobem jako vyjmutí hran v Harrisově detektoru. Hessova matice je určena diferencemi D , které jsou určeny z Taylorova rozvoje.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.50)$$

Kontrola, že se nejedná o hranu při zvoleném parametru r , se provede ověřením podmínky:

$$\frac{\text{trace}(H)^2}{\det(H)} < \frac{(r+1)^2}{r} \quad (3.51)$$

Plný úhel kolem příznaku je rozdělen na histogram s 36 částmi, kde je pro každou část určena hodnota úhlu a váhová funkce, která vychází z gradientu filtrovaného snímku L v daném směru. Výsledná orientace je určena jako maximum histogramu. Pro orientace, které na histogramu dosahují více než 80% nalezeného maxima, je vytvořen nový příznak na stejné pozici s danou orientací.

Pro nalezenou pozici příznaku s daným přiblížením a orientací je na závěr vypočten deskriptor. Nejprve je okolí příznaku transformováno podle získané orientace tak, aby byly příznaky invariantní vůči rotaci. Okolí 16×16 pixelů pro nalezený příznak je v daném měřítku rozděleno na 16 částí o rozměrech 4×4 . V jednotlivých bodech je určen gradient, který je váhován kruhovým Gaussovským oknem. Pro každou část jsou gradienty spojeny do histogramu s osmi směry. Jednotlivé hodnoty histogramů tvoří parametry deskriptoru, kterých je 128 ($4 \times 4 \times 8$). V některých případech může být použito jiné dělení a vzorkování histogramů. Aby byly příznaky robustní i při změně světelných podmínek, jsou vypočtené vektory ještě normovány.

Při přiřazování příznaků na základě rozdílu vzdáleností deskriptorů lze výrazně redukovat počet špatně přiřazených příznaků vyjmutím příznaku, ke kterému kromě nejbližšího existuje další příznak s nepatrně větší vzdáleností. Práh této vzdálenosti je stanoven tak, že poměr minimální vzdálenosti nejbližšího příznaku a dalšího blízkého příznaku vůči přiřazovanému příznaku je větší než 0,8. Experimentálně bylo zjištěno [18], že tímto postupem dochází k potlačení až 90% špatně přiřazených příznaků, zatímco je vyjmuta pouze okolo 5% korektně přiřazených dvojic.

3.4.4 SURF

Hlavní nevýhodou SIFT příznaků je velká výpočetní náročnost, kterou snižuje z ní odvozená metoda SURF (Speeded-Up Robust Features) [22]. Diference Gaussovské funkce počítaná v příznacích SIFT je používána jako aproximace Laplaciánu Gaussovské funkce. Pro SURF příznaky je tato aproximace prováděna pomocí obdélníkových filtrů, ze kterých lze získat parciální derivace ve směrech x , y a xy značené D_{xx} , D_{yy} a D_{xy} .

3. Vizualní odometrie

Detekce příznaků je založená na determinantu Hessovy matice (3.50), který může být aproximován funkcí parciálních derivací:

$$\det(H_{APROX}) \cong D_{xx}D_{yy} - (0,9 \cdot D_{xy})^2 \quad (3.52)$$

Metoda SURF využívá dělení snímků na oktávy a vrstvy, ve kterých nachází stejně jako metoda SIFT význačné body porovnáním diference s okolními body ve snímku a sousedními body ve snímcích předchozí a následující vrstvy. Při porovnávání jsou hodnoty determinantů váhovány podle použitého rozptylu a rozlišení snímku. Jednotlivé vrstvy jsou vytvářeny přímo ze vstupního snímku používáním obdélníkových filtrů různých rozměrů, které nahrazují rozptyl σ . Například filtr s rozměry 9x9 přibližně odpovídá rozptylu σ s hodnotou 1,2.

Orientace nalezených příznaků je hledána vlnkovou transformací ve směrech x , y aplikovanou na okolí příznaku ve vzdálenosti $6s$, kde s je parametrem dané vrstvy. Na parametru s rovněž závisí velikost vlnky. Nalezené odezvy jsou váhovány Gaussovým filtrem s rozptylem $\sigma=2,5 \cdot s$. Orientace je volena pomocí maximální sumy odezvy, která je počítána v jednotlivých krocích posuvného okna představujícího kruhovou výseč s úhlem 60° ze středu příznaku. Tento krok není používán v odvozené metodě U-SURF, čímž dochází k urychlení celého algoritmu. Udávaný rozsah natočení, ve kterých tato metoda zvládá přiřazovat příznaky bez orientace je $\pm 15^\circ$.

Vytváření deskriptoru probíhá na čtvercové oblasti s rozměry $20s$ okolo středu příznaku, která je rozdělena na 16 čtvercových částí s 25 navzorkovanými body. Pro každý bod oblasti je vypočtena odezva na vlnkovou transformaci s Haarovou vlnkou ve směru osy x a y . Tyto odezvy jsou označovány d_x a d_y . Následně je pro každou oblast vytvořen vektor z odezvy v jednotlivých bodech oblasti.

$$\vec{v} = \left[\sum d_x \quad \sum d_y \quad \sum |d_x| \quad \sum |d_y| \right]^T \quad (3.53)$$

Z těchto vektorů pro všechny oblasti je vytvořen deskriptor obsahující 64 prvků. Existuje ale i rozšířený deskriptor se 128 prvky, ve kterém jsou sumy odezev d_x a $|d_x|$ počítány zvlášť pro záporné a nezáporné hodnoty d_y . Také hodnoty součtů d_y jsou počítány zvlášť pro záporná a nezáporná d_x .

Při přiřazování příznaků je možné využít pro urychlení znaménko odpovídající $\text{trace}(H)$, protože tato hodnota udává, jestli je příznak lokálním minimem nebo maximem intenzity. Jednoduchým porovnáním těchto hodnot lze ověřit, zdali si příznaky mohou odpovídat, nebo jsou opačnými extrémy.

3.4.5 FAST detektor

Dalším z detektorů rohů je FAST (Features from Accelerated Segment Test) [23], který vyniká oproti předchozím metodám rychlostí detekce příznaků. Základním principem tohoto detektoru je prohledávání pixelů na kružnici okolo testovaného bodu. Bod p je vyhodnocen jako roh, pokud pro určený počet n bodů v kružnici za sebou platí, že intenzita jednotlivých bodů I_k je splňuje jednu z následujících podmínek:

$$I_p + t < I_k, \quad I_p - t > I_k \quad (3.54)$$

kde t je zvolený parametr. Tyto podmínky vyjadřují, že alespoň n bodů za sebou je výrazněji tmavší nebo světlejší než testovaný bod.

Jedním z možných vylepšení je testování pouze vybraných bodů před samotným testováním všech bodů v kružnici okolo bodu. Například pokud je velikost kružnice okolo bodu $16 \times n=12$, potom stačí otestovat pouze 4 body rovnoměrně rozložené v kružnici. Pokud pro alespoň 3 body neplatí jedna z podmínek (3.54), může být testovaný bod rovnou zamítnut. Nevýhodou testu v této podobě podle autorů článku [23] je, že nelze provádět pro všechny zvolené hodnoty n a ohodnocení jednotlivých pixelů je při testování celé kružnice počítáno znovu.

FAST detektor lze aproximovat rozhodovacím stromem, který je trénován pomocí metod strojového učení. Na základě trénovacích dat s rohy a prázdnými oblastmi je naučen strom, který rozhoduje podle hodnot v jednotlivých pixelech kružnice o daném stavu. Data pro učení mohou být získána analýzou skupiny snímků původním FAST detektorem. Aproximace s rozhodovacím stromem nahrazuje zrychlené testování pro libovolné n a napravuje zahazování hodnot vypočtených v testu. Pokud jsou rohy v trénovací množině určeny, není potřeba znát dopředu přesné parametry pro určování rohů.

Nevýhodou základního FAST detektoru je, že každý bod ze skupiny sousedících pixelů, které náleží stejnému rohu, vyhodnotí jako samostatný příznak. Tuto vlastnost potlačuje metoda Non-maximal Suppression, která okolní příznaky ohodnotí funkcí V . Na základě hodnoty funkce V odstraní ze sousedících příznaků příznaky s nižším ohodnocením. Funkci V lze určit několika různými způsoby. Nejčastěji používaným tvarem funkce V je součet absolutních hodnot rozdílů intenzity pixelů v n bodech oblouku kružnice s bodem p .

$$V(p) = \sum_{k=1}^n |I_k - I_p| - t \quad (3.55)$$

3.4.6 BRIEF deskriptor

Zatím byly v této práci představeny deskriptory pro příznaky SIFT a SURF. Pokud deskriptor těchto příznaků obsahuje 128 hodnot, které jsou v paměti reprezentovány 4 byty, je celková velikost příznaku v paměti počítače 512byte. Navíc určování vzdálenosti na základě Euklidovské vzdálenosti vyžaduje vyšší výpočetní výkon. Obě tyto vlastnosti jsou vylepšeny u BRIEF deskriptoru (Binary Robust Independent Elementary Features) [24], který je reprezentován binárním vektorem. Vzdálenosti deskriptorů je možné určovat pomocí Hammingovy vzdálenosti, která udává počet prvků, ve kterých se binární vektory od sebe liší. Nevýhodou BRIEF deskriptoru je, že není invariantní vůči rotaci. Citlivost deskriptoru BRIEF na rotaci je podobná jako u metody U-SURF, kdy zvládá přiřazovat příznaky rotované až o 15° .

Deskriptor je určován testovací funkcí τ s binárním výstupem. V této funkci jsou porovnávány intenzity dvou oblastí \vec{x} a \vec{y} v okolí příznaku zadané velikosti.

$$\tau(x, y) = \begin{cases} 1, & I_x < I_y \\ 0, & I_x \geq I_y \end{cases} \quad (3.56)$$

Aby nebyl deskriptor ovlivňován šumem, počítá se funkce τ na oblasti snímku, která je filtrována pomocí Gaussovy funkce. Výsledný deskriptor se určí z výsledků funkce τ pro n_d dvojic x, y . Výpočet deskriptoru je popsán následující rovnicí:

$$desc = \sum_{k=1}^{n_d} 2^{k-1} \cdot \tau(x_k, y_k) \quad (3.57)$$

3. Vizualní odometrie

Pro výběr dvojic bodů, ze kterých bude počítán deskriptor, lze použít několik různých přístupů. Body mohou být vybírány tak, aby byly v prohledávané oblasti rozloženy rovnoměrně nebo podle normálního rozdělení. Další možností je použití náhodně vybraných bodů umístěných v diskretizovaných polárních souřadících nebo umístění bodu x do počátku a bodů y na kružnici se středem v počátku. Podle článku [24] je z těchto metod nejlepší metoda s normálním rozdělením bodů.

3.4.7 ORB

Poslední představenou metodou pro detekci příznaků a vytváření deskriptorů bude ORB (Oriented FAST and Rotated BRIEF) [25]. Tato metoda vychází z FAST detektoru příznaků a BRIEF deskriptoru, přičemž upravuje jejich vlastnosti, aby mohla konkurovat robustnějším metodám SURF a SIFT. Oproti těmto metodám je méně výpočetně náročná a není zatížena licenčními omezeními. Nejdůležitější přidanou vlastností je invariance příznaků vůči rotaci. Zajímavostí u této metody je, že byla vyvinuta přímo tvůrci knihovny OpenCV.

Základem detektoru je metoda FAST (kapitola 3.4.5). Aby byly příznaky invariantní vůči zvětšení, jsou detekovány na různě škálovaných snímcích. Dále jsou příznaky ohodnoceny parametrem R z Harrisova detektoru (kapitola 3.4.1), na základě kterého je možné vybrat požadovaný počet nejsilnějších příznaků.

Orientace příznaku je určena těžištěm intenzity, pro které jsou zavedeny proměnné m_{00} , m_{10} a m_{01} . Jejich hodnoty jsou určovány z jednotlivých bodů ležících v kruhu okolo příznaku. Souřadnice těchto bodů jsou značeny x , y a jejich počátek leží ve středu příznaku.

$$m_{00} = \sum_{x,y} I(x,y), \quad m_{10} = \sum_{x,y} x \cdot I(x,y), \quad m_{01} = \sum_{x,y} y \cdot I(x,y) \quad (3.58)$$

Z těchto proměnných lze určit těžiště rovnicí:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (3.59)$$

Orientace příznaku je určena vektorem ze středu příznaku do těžiště intenzity a úhel natočení θ lze určit přímo z proměnných m_{01} a m_{10} .

$$\theta = \text{atan2}(m_{01}, m_{10}) \quad (3.60)$$

Ze získané orientace je možné vytvořit rotační matici pro transformaci množiny bodů binárního testu BRIEF deskriptoru. Aby nebyla tato transformace počítána v každém kroku, jsou pro úhly navzorkované po 12° předpočítány transformované testovací body.

BRIEF deskriptor bez rotace má v jednotlivých bodech velkou směrodatnou odchylku a průměrnou hodnotu blízko 0,5. Další důležitou vlastností je, aby jednotlivé dvojice z binárního testu byly nekorelované. Při zavedení rotace příznaků však dochází ke zhoršení obou těchto vlastností. Jejich zlepšení lze provést pomocí metod strojového učení, kdy jsou pro testované příznaky vybírány dvojice pro vytváření deskriptoru pomocí hladového algoritmu. Nejprve jsou jednotlivé testovací dvojice seřazeny podle vzdálenosti průměrné hodnoty od 0,5. Nejprve je vybrán prvek s nejlepší hodnotou. Pro další prvky je v nalezeném pořadí testováno, jestli s tímto bodem nekoreluje. Pokud ano, tak jsou z testu vyjmuty. V opačném případě je prvek přidán do testovací množiny a hledání pokračuje, dokud není nalezen požadovaný počet bodů. Korelace je vždy testována vůči všem již nalezeným dvojicím.

3.5 Přiřazování a filtrace příznaků

3.5.1 Přiřazování příznaků

Hledání korespondujících bodů mezi dvěma snímky probíhá na základě přiřazování příznaků podle podobnosti jejich deskriptorů. Podobnost deskriptorů je určována jako vzdálenost deskriptorů reprezentovaných vektory. U deskriptorů s číselnými hodnotami, jako jsou deskriptory pro metody SIFT a SURF, je vzdálenost deskriptorů euklidovskou vzdáleností příznakových vektorů. U binárních deskriptorů se naopak využívá Hammingovy vzdálenosti.

Základní metodou pro přiřazování je přiřazení nejbližšího příznak z druhého snímku. Tento přístup je velmi náchylný na špatná přiřazení a může docházet k situacím, kdy několika různým příznakům je přiřazen stejný příznak ve druhém snímku. Také pokud jsou přiřazovány příznaky z prvního snímku do druhého, může být výsledek odlišný od přiřazování v opačném směru.

Přiřazování na základě nejbližší vzdálenosti může být vylepšeno přiřazením příznaků v obou směrech a vybráním pouze korespondencí stejných v obou směrech. Tyto korespondující si příznaky jsou tedy vzájemně svými nejbližšími příznaky. Tím dochází k vyfiltrování slabších příznaků a výraznému omezení špatných přiřazení. Tato metoda bývá označována jako *crossmatching*. Její nevýhodou oproti běžnému přiřazování je delší doba výpočtu.

3.5.1.1 FLANN

Metody založené na vzdálenosti příznaků jsou při velkém počtu příznaků příliš pomalé, a proto někdy bývají nahrazovány různými aproximacemi. Jedním z možných přístupů je rozdělení příznaků z jednoho snímku do stromové struktury, která může být následně pro příznaky z druhého snímku zrychleně prohledávána. Sada metod postavených na prohledávání stromů je představena v článku [26]. Zároveň jsou tyto metody implementovány v knihovně FLANN (fast library for approximate nearest neighbors), která je obsažena i v knihovně OpenCV. Pro příznaky s reálnými hodnotami vektorů jsou představeny metody *priority search k-means tree* a *randomized k-d tree*. Pro binární vektory slouží metoda *hierarchical clustering tree*.

Metoda *priority search k-means tree* je založena na shlukování příznaků. V každém kroku jsou deskriptory rozděleny do zadaného počtu shluků k pomocí k -means algoritmu, které jsou určeny svým těžištěm a tvoří uzly stromu. Každý shluk je dále dělen do k skupin, dokud skupina neobsahuje méně než k prvků. Při prohledávání jsou při průchodu k nejbližšímu listu stromu zaznamenávány do fronty další nejbližší nalezené uzly pro další případná prohledávání stromu.

Metoda *randomized k-d tree* využívá upravené metody k -d stromu, kdy proměnná (dimenze) podle které probíhá dělení je vybírána náhodně ze zvoleného počtu dimenzí s největší odchylkou. Těchto stromů je vytvořeno několik a následně jsou prohledávány paralelně. Ze všech stromů je vždy prohledáván uzel, ze kterého je nejbližší k listu. Pro získání přesnějšího výsledku nemusí být použit první nalezený list, ale výsledek může být vybrán až na základě více nalezených listů.

V metodě *hierarchical clustering tree* pro binární deskriptory je sestrojeno několik různých stromů. Vytváření stromu je podobné první metodě, akorát uzly tvoří přímo náhodně zvolené deskriptory a ostatní body jsou rozděleny do k skupin podle vzdálenosti od takto náhodně zvolených bodů. Každý ze stromů je prohledáván zvlášť. Pokud je vyžadováno více bodů na výstupu, je pro každý strom navíc vytvořena prioritní fronta. Výsledkem jsou nalezená přiřazení s nejnižší vzdáleností.

3.5.2 RANSAC

Mezi přiřazenými příznaky se zpravidla vyskytují i chybná přiřazení, která mohou negativně ovlivnit přesnost dalších výpočtů. Často používanou metodou ve vizuální odometrii pro vyfiltrování vlivu špatně přiřazených příznaků na výsledné řešení je metoda RANSAC ([15], str. 83), která je založena na principu náhodného výběru bodů. Její hlavní výhodou je, že dokáže pracovat i s velkým počtem špatně přiřazených bodů. Naopak za nevýhodu lze označit nedeterministické chování, které je způsobeno náhodným výběrem bodů.

Základním principem této metody ve vizuální odometrii je výpočet modelu pohybu pro několik náhodně zvolených množin bodů, jejichž řešení je vždy ohodnoceno na zbylých bodech. Za výsledné řešení je považován výsledek s nejlepším ohodnocením. Použitými metodami pro výpočet pohybu mohou být například metody představené v kapitole 3.3.

Počet vybraných množin nezbytný pro nalezení správného řešení je určen rovnicí:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - e)^s)} \quad (3.61)$$

p v rovnici představuje požadovanou pravděpodobnost úspěchu algoritmu. s udává počet bodů potřebný pro výpočet modelu pohybu a e představuje podíl špatně přiřazených bodů v celé množině. V reálných aplikacích je zpravidla používán desetkrát větší počet, což zaručuje větší robustnost algoritmu.

Potřebný počet kroků N je exponenciálně závislý na počtu bodů pro výpočet modelu pohybu. Protože velký počet kroků vede k velkému zpomalení celého výpočtu, bývá často cílem použití metody vyžadující pro určení posunu nejmenší možný počet bodů. Počet bodů může být kromě výběru metody s minimálním počtem bodů snížen i omezením počtu stupňů volnosti modelu pohybu.

Naopak pokud jsou pozice bodů výrazně ovlivněny šumem, bývá výhodnější použití metody s větším počtem bodů, u které dochází k většímu potlačení šumu.

3.6 Systémy a metody pro vizuální odometrii

3.6.1 SOFT-SLAM

Jednou z možných aplikací vizuální odometrie je sledování trajektorie automobilu z kamer umístěných na vozidle. Tyto systémy mohou být porovnávány v KITTI benchmarku [27], ve kterém jsou jednotlivé algoritmy testovány na stejných datech s referenčními pozicemi. Kromě dat ze stereo kamery mohou být systémy postaveny na laserovém dálkoměru. Nejlepší metodou v KITTI benchmarku, která je postavena na vizuální odometrii, a čtvrtou nejlepší metodou celkově je SOFT-SLAM [28]. Systém se skládá ze dvou vláken. V prvním vlákne je prováděna vizuální odometrie, která vychází ze starší metody od stejných autorů [29], a druhé vlákno zajišťuje vytváření mapy okolí. Samotná odometrie bez mapování je v KITTI 6. nejlepším systémem postaveným na vizuální odometrii. Navíc v době svého uvedení v roce 2015 byla v KITTI nejlepším systémem s vizuální odometrií.

Vizuální odometrie SOFT-SLAMu je postavena na silné filtraci příznaků. Pro detekci příznaků není používána žádná z metod popsaných v kapitole 3.4. Používané příznaky vychází z článku [30]. Založeny jsou podobně jako metoda SURF na obdélníkových filtrech o velikosti 5x5px. Používány jsou dva různé obdélníkové filtry pro detekci rohů a kontrastních objektů. Příznaky jsou nalezeny jako minima a maxima ve filtrovaných snímcích. Deskriptory jsou

popsány odezvami na Sobelův filtr ve vybraných oblastech v okolí příznaku a přiřazovány podle vzdálenosti deskriptorů. Přiřazování je možné urychlit rozdělením příznaků do čtyř skupin, podle toho, jestli byl detekovaný příznak rohem nebo kontrastní oblastí a minimální nebo maximální hodnotou.

Filtrace příznaků je prováděna v několika krocích, aby byly zachovány pouze silné příznaky. V prvním kroku je používána metoda nazývaná *circular matching*. Příznaky jsou přiřazovány mezi dvěma po sobě následujícími dvojicemi stereo snímků. Posupně jsou hledány korespondence mezi pravým a levým snímkem první dvojice, oběma levými snímky, pravým a levým snímkem druhé dvojice a na závěr mezi pravými snímky. Pokud nalezený příznak v přiřazení mezi pravými snímky neodpovídá počátečnímu příznaku, jsou všechny tyto příznaky vyjmuty z množiny příznaků. V dalším kroku jsou pro zbylé příznaky přiřazení ověřena normalizovanou vzájemnou korelací v okolí příznaku o rozměrech 25x25px. Posledním z kroků při filtraci příznaků je rozdělení snímku na čtvercové oblasti 50x50px, ze kterých je zachován pouze určený počet nejsilnějších příznaků. Zbylá chybná přiřazení jsou ošetřena použitím RANSAC algoritmu.

Další vlastností příznaků je jeho sledování mezi snímky v čase. Každý příznak má určeno své stáří, kterým je hodnota, v kolika snímcích již byl detekován. Stáří příznaku je používáno jako hlavní parametr jeho síly. Při sledování příznaku je používán jeho počáteční deskriptor a pozice v následujících snímcích je zpřesňována na základě znalosti pozice ve snímcích předchozích.

Pro výpočet posunu z nalezených příznaků je od sebe oddělen výpočet rotace a translace mezi snímky. Pro určení rotace je používána pětibodová metoda (kapitola 3.3.2.2), která je spojena s RANSAC algoritmem. Výsledné řešení je vybráno na základě esenciální matice určené z matice rotace a translace vrácené pětibodovou metodou, která vrátí nejméně chybných přiřazení. Použití pětibodové metody, která vyžaduje nejmenší počet bodů z metod 2D do 2D, snižuje pravděpodobnost, že zvolená množina bude obsahovat chybnou hodnotu. Výpočet rotace pouze z jedné kamery odstraní chybu způsobenou nepřesností kalibrace stereopáru. Dále vyžaduje přiřazení mezi dvěma snímky namísto čtyř, čímž současně snižuje pravděpodobnost špatného přiřazení příznaků. Při sledování příznaků mezi snímky může být výsledná rotace mezi snímky v čase i a $i-1$ dále zpřesněna za použití přiřazení i mezi snímky i a $i-2$. Samotná translace je počítána metodou 3D do 2D. Triangulované body z předchozího snímku jsou zpětně promítány do pravého i levého aktuálního snímku. Výsledné řešení je hledáno iterativní metodou, která hledá translaci tak, aby byla minimalizována odchylka zpětně promítnutých bodů od naměřených bodů v obou snímcích.

U obou systémů navíc autoři počítají s jeho možným vylepšením na základě dat z jiných senzorů a následným odhadem pozice pomocí Kalmanova filtru. Mapovací vlákno detekuje uzavírání smyček a zajišťuje globální optimalizaci nalezené trajektorie.

3.6.2 RotRocc++

Další z metod 3D do 2D, která dosahuje velmi dobrých výsledků v KITTI benchmarku (4. nejlepší systém postavený na vizuální odometrii), je RotRocc++ [31]. Obdobně jako SOFT-SLAM využívá odděleného výpočtu rotace a translace. Pro detekci příznaků používá metodu Shi-Tomasi. Příznaky přiřazené na základě minimální vzdálenosti trianguluje čase $i-1$. K odstranění chybných příznaků využívá upravené metody *circular matchingu*.

Pro nalezené příznaky je nejprve iterativně určována rotace. K jejímu určení je ale potřeba určovat i translaci. Jako počáteční hodnoty jsou používány transformace nalezené pro

3. Vizualní odometrie

předchozí snímky. Z nalezeného posunu v každé iteraci jsou filtrovány špatně přiřazené body, jejichž rozdíl naměřené a zpětně promítnuté pozice překračuje zadanou hranici. Tato hranice je určena jako maximální hodnota z pevně stanovené hranice a n -tého nejmenšího rozdílu. Nalezená rotace je použita pro transformaci příznaků ze snímků v čase i . V dalším kroku je určována pouze translace.

Hledání translace probíhá opět iterativně s filtrováním příznaků. Při filtrování je používána pro určení odchylky zpětně promítnutého a naměřeného příznaku oddělená normalizovaná odchylka zpětné projekce (decoupled normalized reprojection error).

$$\delta_n^i = \frac{\varepsilon_n^i(R^i, \vec{t}^i)}{\|\vec{x}_n^{i-1} - \pi(R^i \vec{x}_n^i)\|} \quad (3.62)$$

ε je odchylka zpětné projekce pro nalezenou transformaci od naměřené a π je transformace bodu \vec{x} v obraze podle rotace R .

Výsledná transformace je spojením nalezené translace s rotací určenou v předchozím kroku. Pro malé rychlosti vozidla je translace určována již z prvního kroku, aby nebyl výsledek ovlivněn numerickými nepřesnostmi, které způsobí malé změny v obraze.

Důležitou částí RootRoc++ metody je zjednodušení rotační matice, ve které jsou předpokládány pouze malé změny v úhlu do 5° . Pro takto malé úhly je možné aproximovat goniometrické funkce v rotační matici Taylorovým polynomem prvního řádu.

3.6.3 Monokulární vizualní odometrie pro silniční vozidla

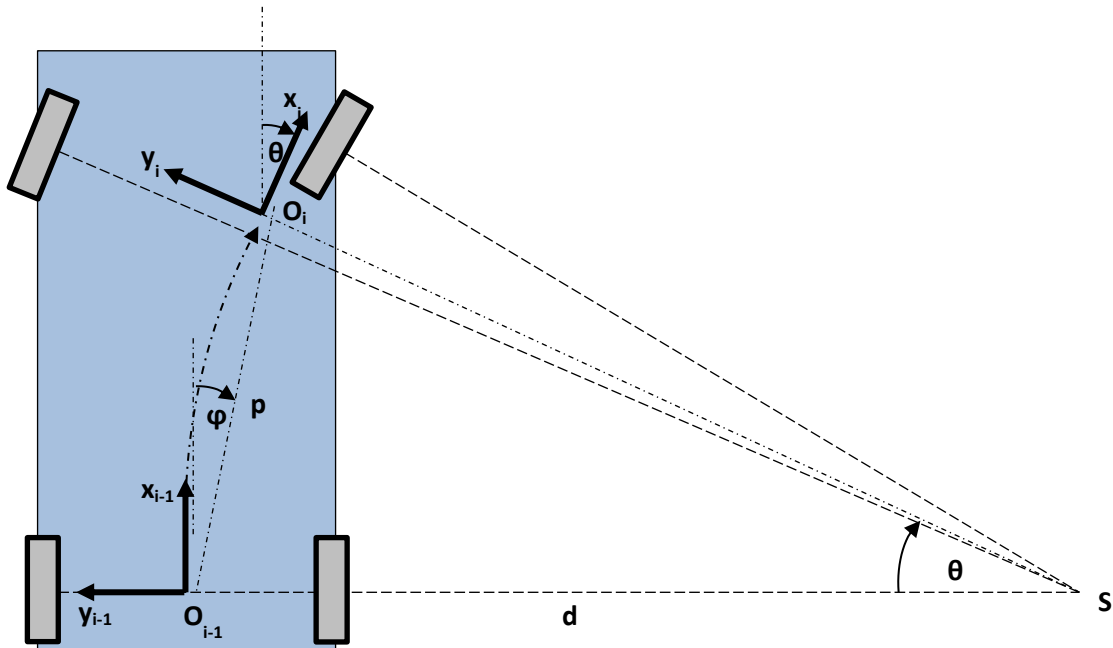
Obecně je pro přesné určení transformace mezi dvěma pozicemi pevného tělesa ve 3D prostoru zapotřebí 6 parametrů. Pokud je však popisován pohyb automobilu, který je omezen geometrií podvozku, je možné pro pohyb snížit počet stupňů volnosti. Extrémním případem zjednodušení je metoda pro monokulární vizualní odometrii představená v článku [32], která pro vyjádření pohybu používá pouze jeden parametr.

Základním předpokladem pro omezení pohybu je geometrie Ackermannova podvozku, u kterého vozidlo opisuje kruhový pohyb podle kružnice se středem v průsečíku os jednotlivých kol (Obr. 8). Dále lze pro pohyb automobilu předpokládat pouze rovinný pohyb, pro jehož popis stačí pouze 3 parametry. Dva parametry obvykle představují souřadnice v osách x , y a třetím parametrem je úhel natočení θ . Vzhledem ke kruhovému pohybu Ackermannova podvozku je možné nahradit souřadnice x , y polárními souřadnicemi p , φ . Na základě součtu úhlů v trojúhelníku $\Delta O_{i-1} S O_i$ je možné odvodit vztah, mezi novou souřadnicí φ a rotací θ mezi souřadnými soustavami v čase i a $i-1$.

$$\varphi = \frac{\theta}{2} \quad (3.63)$$

Tím je popis pohybu zredukován na pouhé dva parametry. Pro monokulární metody navíc není možné přesně určit měřítko získané trajektorie, které představuje jeden ze stupňů volnosti. Pro popis pohybu automobilu proto stačí pouze jeden parametr. Výsledná rotace a translace je popsána následujícími vztahy.

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad t = p \cdot \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \\ 0 \end{bmatrix} \quad (3.64)$$



Obr. 8 Ackermannův podvozek a znázornění pohybu po kružnici

Jediným parametrem popisujícím pohyb je úhel θ , který je možné určit z esenciální matice E . Aby byly souřadnice bodů nezávislé na modelu kamery, jsou body v souřadnicích kamery promítnuty na sféru s poloměrem 1. Po aplikování epipolární podmínky platí pro úhel θ vztah:

$$\theta = -2 \cdot \arctan\left(\frac{y_i z_{i-1} - z_i y_{i-1}}{x_i z_{i-1} + z_i x_{i-1}}\right) \quad (3.65)$$

Dalším možným přístupem je řešení epipolární podmínky pro více získaných bodů metodou SVD s neznámými $\sin\left(\frac{\theta}{2}\right)$ a $\cos\left(\frac{\theta}{2}\right)$.

Předchozí popsané vztahy platí pouze pro kameru, jejíž střed promítání je umístěn nad osou zadních kol. Toto omezení však není zpravidla nutné dodržovat vzhledem k velkým poloměrům kružnice pohybu.

Na základě omezení pohybu na jeden parametr navrhli autoři článku [32] dvě metody filtrace špatných přiřazení. První metodou je jednobodový RANSAC. Počet iterací v RANSACu je ovlivněn počtem chybných hodnot a počtem bodů v jednom výběru. Vzhledem k exponenciální závislosti počtu iterací na počtu bodů ve výběru dochází při jednobodové metodě k výraznému snížení počtu iterací. Druhou metodou je filtrování na základě histogramu, kdy jsou nalezené úhly rozděleny do malých intervalů a zvolen je interval obsahující nejvíce hodnot. Pro zjednodušení výpočtu je možné nahradit histogram výpočtem mediánu úhlu θ .

Hlavní výhodou této metody jsou nízké nároky na výpočetní výkon a počet příznaků. Dále minimální počet hodnot pro výpočet snižuje pravděpodobnost ovlivnění výsledku chybnou hodnotou. Metoda je také robustní vůči mnoha jiným objektům, které se pohybují v obraze nezávisle na vozidle. Navíc je u této metody možné použití kamer s jiným modelem, než je model perspektivní kamery.

3.6.4 Zjednodušení epipolární geometrie pro kolová vozidla

Pro omezení počtu stupňů volnosti u automobilů existuje více možných přístupů. Přístup popsán v předchozí kapitole může být až příliš restriktivní a negativně ovlivňovat přesnost určení polohy. Tuto skutečnost na několika různých datasetech pro kolová vozidla analyzovali autoři v článku [33] a představují 2 vlastní metody, které jsou méně restriktivní. Výhodou metod s omezením počtu stupňů volnosti pohybu je především zrychlení výpočtu posunu.

Prvním zjednodušeným modelem je model se 4 stupni volnosti, který vychází z modelu pro pohyb v rovině se třemi stupni volnosti. Model se 3DoF je určen dvěma úhly a poloměrem kružnice reprezentujícím měřítko. Matice rotace a translace pro 3DoF systém mohou být vyjádřeny následujícími rovnicemi:

$$R_{3D} = \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) \\ 0 & 1 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix}, \quad t_{3D} = d \cdot \begin{bmatrix} \cos(\varphi_{xz}) \\ 0 \\ \sin(\varphi_{xz}) \end{bmatrix} \quad (3.66)$$

Souřadný systém je uvažován s osou z ve směru osy kamery a osou x rovnoběžnou s rovinou pohybu. Obě rotace proto probíhají ve směru osy y. Rozšiřující stupeň volnosti slouží pro zahrnutí vibrací vozu, které jsou reprezentovány jako pohyb podél osy y. Tento pohyb může být do modelu zahrnut buď jako posun ve směru y nebo jako rotace podél osy x. Rotační a translační matice mohou být pro rozšířený model podle způsobu reprezentace posunu ve směru osy y vyjádřeny v následujících dvou tvarech:

$$R_{4D} = R_{3D} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) \\ 0 & \sin(\theta_x) & \cos(\theta_x) \end{bmatrix}, \quad t_{4D} = t_{3D} \quad (3.67)$$

$$R_{4D} = R_{3D}, \quad t_{4D} = d \cdot \begin{bmatrix} \cos(\varphi_{yz}) \cdot \cos(\varphi_{xz}) \\ \sin(\varphi_{yz}) \\ \cos(\varphi_{yz}) \cdot \sin(\varphi_{xz}) \end{bmatrix} \quad (3.68)$$

Pro odhad posunu bez měřítka modelu s 4DoF stačí pouze tři parametry. Pro výpočet na základě epipolární matice používají autoři článku Newtonovu metodu.

Druhým představeným zjednodušeným modelem pohybu je upravený 6DoF model. Pro kolová vozidla vlivem geometrie Ackermannova podvozku a téměř rovinným pohybem, je jediným úhlem, který se výrazněji mění, rotace podle osy y. Ostatní úhly se mění jen nepatrně zpravidla vlivem různých nerovností na vozovce a vibracemi vozidla. Proto mohou být rotace podle os x a z aproximovány Taylorovým polynomem 1. řádu. Takto upravená matice rotace má následující tvar, kde je vektor translace vyjádřen v kartézských souřadnicích.

$$R = \begin{bmatrix} -\cos(\theta_y) & -\theta_z & \sin(\theta_y) \\ \theta_z \cdot \cos(\theta_y) & 1 & -\theta_x \\ -\sin(\theta_y) & \theta_x \cdot \cos(\theta_y) & \cos(\theta_y) \end{bmatrix}, \quad t = d \cdot \begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} \quad (3.69)$$

Pro monokulární kameru vede tento případ na aproximované řešení pětibodové metody. Tu autoři [33] řeší opět pomocí Newtonovy metody, čímž se vyhýbají řešení polynomu 10. řádu v pětibodové metodě. Tím dochází k výraznému urychlení algoritmu. Přestože není snížen počet proměnných, dochází ke zrychlení RANSAC algoritmu zrychlením dílčích výpočtů jednotlivých odhadů.

3.6.5 Metoda pro robustní odhad pohybu

Další z možných přístupů k vizuální odometrii je popsán v článku [34]. Autoři přichází s myšlenkou využití všech kanálů RGB obrazu, zatímco většina ostatních systémů pracuje pouze s černobílým obrazem. Protože zpravidla použité kamery zachycují barevný obraz, dochází při převodu ke ztrátě informace. Příznaky jsou v představeném systému detekovány nezávisle v každém ze tří kanálů RGB snímku a následně spojeny do příznaku s přesněji určenou polohou. Další částí této metody je vyjádření kovariančních matic jednotlivých příznaků, které v sobě zachycují nejistotu určení polohy nalezeného příznaku. Tyto matice jsou využívány nejen při slučování příznaků z jednotlivých kanálů, ale i při přiřazování příznaků mezi snímky.

Pro detekci příznaků je používán Harrisův a SIFT detektor. Výpočet Harrisových příznaků je založen na určování matice M (kapitola 3.4.1), ze které lze odhadnout kovarianční matici reprezentující nejistotu příznaku. Tuto matici lze určit přímo jako inverzní matici k matici M . U SIFT příznaků je místo matice M použita Hessova matice, ze které lze kovarianční matici získat rovněž její inverzí. Spojení kovariančních matic z jednotlivých kanálů je váhovaným součtem nalezených matic.

Kovarianční matice může zároveň reprezentovat elipsu kolem příznaku, která udává nejistotu jeho polohy. Tuto informaci lze použít pro filtraci špatně přiřazených příznaků, kdy je správné přiřazení detekováno jako průnik elips zpětně promítnutého a naměřeného příznaku. Kovarianční matici lze použít i při výpočtu fundamentální matice reprezentující pohyb mezi snímky k určení váhy jednotlivých příznaků. Váhování příznaků zaručí menší vliv příznaků s velkou nejistotou na konečný výsledek.

Využití informace ze všech kanálů vede k přesnějším výsledkům než použití pouze černobílého obrazu. Jeho nevýhodou je naopak nutnost výpočtu příznaků ve všech kanálech, což negativně ovlivní čas výpočtu.

3.6.6 SVO

Posledním představeným systémem pro vizuální odometrii bude SVO (Semi-Direct Visual Odometry) [35], který je druhou verzí systému představeného v roce 2014 [36]. Původní systém byl navržen čistě pro jednu kameru, zatímco novější verze umožňuje využití více kamer. Hlavním přínosem použití více kamer je možnost triangulace bodů při spuštění algoritmu a správná inicializace počátečních poloh bodů v mapě. U původní verze jsou přístupné zdrojové kódy systému, zatímco v novější verzi jsou autory uvolněny pouze binární soubory s licencí pro nekomerční použití.

SVO provádí zároveň vytváření mapy bodů, která umožňuje optimalizaci v případě uzavření smyčky a přesnější lokalizaci. Aby vytváření mapy neovlivňovalo rychlost, je odděleno od lokalizace do samostatného vlákna. Přestože metoda je navržena pro monokulární systémy, používá pro výpočet posunu mezi snímky metodu „3D do 2D“. Využití této metody umožňují známé odhady hloubky bodů v předchozím snímku, které jsou získávány z mapy příznaků. Při hledání posunu mezi snímky nejsou příznaky ve snímku popsány deskriptorem, ale přiřazování probíhá na základě intenzity jednotlivých pixelů mezi transformovaným předchozím a aktuálním snímkem.

Protože je tento posun odhadnut pouze na základě předchozího snímku, docházelo by ke sčítání chyb jednotlivých transformací. Aby bylo tomuto jevu zabráněno, je nalezené řešení dále zpřesněno pomocí zpětné projekce všech 3D bodů z mapy, které by měly být zachyceny v aktuálním snímku a minimalizace chyby jejich zpětně promítnutých pozic od změřených.

3. Vizualní odometrie

Hlavní úlohou mapovacího vlákna je vytváření nových 3D bodů v mapě. Pro každý bod je kromě jeho 3D pozice určena i nejistota odhadu hloubky. 3D bod je vždy přiřazen konkrétnímu snímku, ve kterém je určena hloubka tohoto bodu. Při hledání korespondujícího bodu k příznaku je snímek prohledáván podél epipolární přímky a hledán bod, jehož okolí nejvíce koreluje s okolím příznaku. Nejistota určení hloubky příznaku je snižována dalšími pozorováními tohoto příznaku. Při použití pouze jedné kamery jsou pozice určené s nejistotou v měřítku, která může být odstraněna použitím dalšího externího senzoru nebo počáteční inicializací se známými délkami posunu mezi snímky.

Aby nedocházelo ke zpomalování systému při dlouhém běhu, je předem omezen počet snímků ukládaných v mapě. Přidávány jsou pouze snímky s výrazně větší vzdáleností od ostatních snímků. Pokud při přidání snímku dojde k překročení počtu snímků, je vyjmut snímek, který se nachází nejdále od nově vloženého. Pro hledání příznaků je používán FAST detektor, který je aplikován na pyramidu převzorkovaných snímků o pěti vrstvách. Kromě rohů systém využívá i detekci hran.

Tento systém vyniká zejména nízkými nároky na výkon. Primárně je určen pro drony, které bývají osazeny méně výkonnými výpočetními jednotkami. Přesto na těchto jednotkách dokáže v reálném čase určovat polohu rychlostí až 50 snímků za sekundu pro kameru s rozlišením 752x480px.

4 Navržené metody

4.1 Metoda pro automatické testování vlastností příznaků

Pro výběr metody bylo klíčové najít správný typ příznaku, aby bylo možné detekovat ve všech snímcích podvozku dostatečný počet příznaků, které zároveň půjde mezi snímky správně přiřadit. Vzhledem k velkému počtu parametrů pro jednotlivé metody detekce příznaků a vytváření jejich deskriptorů, které významně ovlivňují vlastnosti používané metody, byla v této práci navržena metoda pro jejich automatické testování.

Pro každý parametr byla zvolena množina testovaných hodnot. V automatickém testu byly následně na sadě testovacích stereo snímků podvozku testovány jednotlivé kombinace nastavení všech parametrů. V každém stereo snímku byly detekovány a přiřazeny příznaky. Přiřazování probíhalo na základě hledání nejbližšího deskriptoru za použití *crossmatchingu*. Při detekci byl snímek rozdělen na 12 částí (4x3), ve kterých probíhala detekce příznaků zvlášť. Toto dělení bylo použito i pro vizuální odometrii, která je popsána v kapitole 4.2.

Každá přiřazená dvojice byla otestována, jestli splňuje epipolární podmínku. Epipolární podmínka byla považována za splněnou, pokud vzdálenost příznaku od epipolární přímky spočtené z polohy příznaku v druhém snímku nepřekročila ani v jednom snímku zadaný práh. Parametry přímky byly získány z fundamentální matice F podle rovnice (3.5), jejímž výstupem je třírozměrný vektor $\vec{l} = [l_1, l_2, l_3]^T$ obsahující jednotlivé koeficienty rovnice přímky. Vzdálenost bodu s obrazovými souřadnicemi u, v od přímky lze určit z rovnice:

$$d = \frac{l_1 u + l_2 v + l_3}{\sqrt{l_1^2 + l_2^2}} \quad (4.1)$$

Pro přiřazené body splňující tuto podmínku byl triangulací (kapitola 3.2.2.2) určen 3D bod v prostoru. Na základě známé informace o přibližné vzdálenosti podvozku od kamery byla testována souřadnice z nalezených 3D bodů, jestli leží v zadaném okolí známé vzdálenosti podvozku. Přiřazení, která splnila obě tyto podmínky, byla označena za správná.

Měřenými parametry pro jednotlivé kombinace byly celkový počet správně přiřazených příznaků, celkový počet přiřazených příznaků a minimální počet správných přiřazení v jednom stereo snímku. Dále byla zaznamenána maximální délka trvání výpočtu příznaků a jejich přiřazení v jednom stereo snímku.

4.2 Metoda pro vizuální odometrii

V této kapitole bude představena metoda 3D do 2D založená na postupech popsaných v kapitole 3. Obdobně jako metoda [29] využívá silnou filtraci příznaků a oddělený výpočet rotace a translace. Pro určení rotace je kromě RANSAC algoritmu používán algoritmus pro řešení problému tří bodů z článku [14]. Translace je počítána z více bodů řešením přeurené soustavy lineárních rovnic metodou nejmenších čtverců. Jednotlivé kroky metody jsou podrobněji popsány dále v této kapitole.

Při návrhu nebyla uvažována možnost optimalizace metody přes více než dva po sobě následující snímky, protože je předpokládáno, že překryv mezi po sobě následujícími snímky bude co největší. Zároveň vzhledem k pohybu vozidla pouze jedním směrem metoda nepočítá s možným uzavíráním smyček, které by umožnily další optimalizaci získaných výsledků. Pro ponechání co největší robustnosti metody nebyl omezován počet stupňů volnosti vozidla.

4.2.1 Detekce a přiřazování příznaků

Vstupem metody je trojice snímků a parametry z kalibrace kamer. První dva snímky jsou snímky ze stereo páru kamer pořízené v čase $i-1$. Třetí snímek pochází z první kamery v čase i . Souřadnice první kamery v čase i jsou považovány za globální prostorové souřadnice, vůči kterým je určována výsledná poloha kamer v předchozím snímku.

Pro detekci a popis příznaků jsou používány metody z knihovny OpenCV. Vzhledem k charakteru podvozku auta při testování detekce na celém snímku v některých případech docházelo k detekci většiny příznaků v místě s velkými změnami intenzity, což vedlo k selhání celé metody, pokud se toto místo již nenacházelo v posunutém snímku. Částečným řešením je nastavení většího počtu příznaků, které ale zvyšuje výpočetní náročnost. Při detekci byly proto snímky rozděleny v horizontálním i vertikálním směru na několik obdélníkových částí, ve kterých probíhala detekce příznaků zvlášť. Toto rozdělení zajišťuje rovnoměrnější rozložení příznaků ve snímku a tím i lepší stabilitu celé metody.

V prvním kroku jsou příznaky detekovány ve všech třech snímcích, ale přiřazovány jsou pouze příznaky mezi prvním a druhým snímkem. Přiřazování vůči třetímu snímku je vhodné provádět až po filtraci příznaků pro snížení výpočetní náročnosti přiřazování i zabránění úbytku příznaků vlivem špatných přiřazení. Přiřazování probíhá na základě hledání nejbližších deskriptorů s *crossmatchingem* (kapitola 3.5.1).

4.2.2 Filtrace příznaků a triangulace

V prvním kroku byly filtrovány příznaky úspěšně přiřazené mezi prvními dvěma snímky. Použitý filtr byl téměř stejný jako filtr pro automatické testování příznaků popsany v kapitole 4.1. Nejprve bylo ověřeno, jestli vzdálenost příznaku ve snímku od epipolární přímky nepřekročí zadanou hranici. Pokud byla splněna tato podmínka, byla provedena triangulace a filtrace na základě vzdálenosti příznaku od kamery v ose z . Protože však nelze předpokládat, že v reálné aplikaci bude známá informace o vzdálenosti podvozku od kamery jako při automatickém testování, byl rozsah vzdáleností určen v pevně v zadaném intervalu.

Další krok filtrování příznaků vychází z metody *circular matchingu* z článku [29] (kapitola 3.6.1), u které je kontrola správného přiřazení prováděna postupně mezi čtyřmi snímky. Upravená metoda bude dále nazývána *triangular matching*, protože provádí kontrolu přiřazení pouze mezi třemi snímky. V tomto kroku jsou provedena zbylá přiřazení mezi snímky 1-3 a 2-3, při kterých jsou ve 3. snímku hledány korespondující příznaky pouze k vyfiltrovaným příznakům ze snímků 1 a 2. Při *triangular matchingu* je nejprve vybrán příznak z přiřazení mezi snímky 1-3, k jehož pozici ve třetím snímku je hledáno odpovídající přiřazení mezi snímky 2-3. Pokud toto přiřazení není nalezeno, dojde k vyjmutí tohoto příznaku. Pokud existuje, je provedena kontrola, že příznak ve snímku 1 odpovídá příznaku ve snímku 2 podle přiřazení 1-2. Tuto kontrolu usnadňuje, pokud jsou vyfiltrované příznaky ve snímcích 1 a 2 seřazeny tak, aby jejich pozice v poli příznaků ze snímků 1 a 2 byly stejné pro korespondující příznaky. Pro seřazené příznaky následně stačí ověřit, že si pozice příznaků, které jsou výsledkem přiřazení mezi snímky 1-3 a 2-3, v polích příznaků odpovídají.

Posledním krokem filtrace je vyjmutí příliš blízkých příznaků. Při kontrole jsou procházeny příznaky pouze v prvním snímku, pro které jsou v okolí hledány nejbližší příznaky. Pokud jsou v zadané vzdálenosti od příznaku nalezeny další příznaky, je ponechán pouze příznak s nejlépe ohodnoceným deskriptorem.

4.2.3 RANSAC a výpočet rotace

Pro výpočet rotace mezi snímky byl použit RANSAC algoritmus popsany v kapitole 3.5.2. Z přechozích kroků je získána množina příznaků, pro které je kromě jejich pozice ve všech snímcích známa i jejich triangulovaná poloha ve 3D v čase $i-1$. Na začátku algoritmu jsou náhodně vybrány 4 příznaky, pro které bude řešena úloha tří bodů metodou z článku [14]. Vstupem této metody jsou 3D souřadnice tří bodů a normované vektory směřující k příznakům v souřadnicích kamery v čase i . Obrazové souřadnice příznaků ze snímku 3 je proto potřeba převést do souřadného systému kamery vynásobením inverzní maticí k matici kamery a získaný vektor vydělit jeho velikostí, aby délka vektoru byla 1. Výstupem P3P metody jsou 4 možná řešení. Každé řešení je reprezentováno rotační maticí R a vektorem translace \vec{t} . Vracené řešení je ale transformací souřadného systému kamery do prostorových souřadnic 3D bodů. Opačnou transformaci, která je vyžadována popisovanou metodou lze získat převodními vztahy:

$$R_v = R^T, \quad \vec{t}_v = -R_v \cdot \vec{t}_v \quad (4.2)$$

Výběr správného řešení bude proveden pomocí čtvrtého bodu, pro který se vypočte rozdíl mezi zpětně promítnutým bodem podle nalezené transformace a jeho skutečným obrazem. Řešení s nejmenší odchylkou bude vráceno jako správné. Zpětný průmět transformovaného bodu \vec{u}_4 lze spočítat z následujících rovnic, kde \vec{X}_4 představuje 3D bod získaný triangulací mezi snímky 1-2:

$$\vec{u}_{4d} = K_1 \cdot (R_v \cdot \vec{X}_4 + \vec{t}_v), \quad \vec{u}_4 = \frac{\vec{u}_{4d}}{u_{3,4d}} \quad (4.3)$$

Chybu zpětné projekce lze pro změřené obrazové souřadnice \vec{u}_{4m} lze vyjádřit jako:

$$e = \|\vec{u}_4 - \vec{u}_{4m}\| \quad (4.4)$$

Výsledek získaný ze čtyř náhodně zvolených bodů je následně ověřen na celé množině filtrovaných příznaků. Nalezené řešení bylo ohodnocováno na základě chyby zpětné projekce. Ohodnocení počítané přes všechny body bylo zvoleno jako počet bodů, pro které chyba zpětné projekce nepřekročí zadaný práh t_h . Ohodnocovací funkci H lze vyjádřit následujícími rovnicemi, ve kterých $h(k)$ představuje ohodnocení pro k -tý testovaný bod:

$$h(k) = \begin{cases} 1, & e_k \leq t_h \\ 0, & e_k > t_h \end{cases} \quad (4.5)$$

$$H = \sum_k h(k) \quad (4.6)$$

Za výslednou transformaci vrácenou RANSAC algoritmem je zvoleno řešení, které má nejvyšší hodnotu ohodnocovací funkce H .

Výpočetní náročnost tohoto algoritmu je přímo úměrná počtu kroků algoritmu, který je závislý na velikosti podílu chybných dat (kapitola 3.5.2), a není možné ho snižovat pod určitou mez. Při použití procesoru s více jádry je možné výpočet rozdělit do více vláken. Vzhledem k náhodnému výběru bodů lze spustit několik RANSAC algoritmů současně s menším počtem iterací, který je nepřímo úměrný počtu vláken vzhledem k celkovému počtu iterací. Výsledným řešením ze všech vláken bude řešení s nejvyšší hodnotou ohodnocovací funkce H .

4.2.4 Výpočet translace

Výpočtu translace předchází dvě filtrování příznaků. V prvním kroku jsou zachovány pouze příznaky splňující podmínku z rovnice (4.5). Druhým krokem je mediánový filtr, který seřadí zbylé příznaky na základě velikosti jejich odchylky při zpětné projekci z rovnice (4.4) a vypočte jejich medián. Pro další výpočet vybere pouze polovinu bodů, které jsou nejbližší mediánu. Tento filtr byl navrhován experimentálně. Jeho cílem je snížit počet příznaků pro další výpočet a omezit závislost výsledku na řešení získaném z výpočtu rotace.

Pro zbylé příznaky je řešena přeurčená soustava rovnic, ve které neznámou představují pouze prvky vektoru translace \vec{t} , které budou značeny t_1, t_2, t_3 . Naměřené pozice bodu ve 3. snímku budou u, v a souřadnice bodu v 3D prostoru vůči snímkům v čase $i-1$ \vec{X} . Vztah obrazu bodu a 3D souřadnic lze vyjádřit vztahem podle rovnice (3.1):

$$\lambda \cdot \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} = K \cdot R \cdot \vec{X}_k + K \cdot \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (4.7)$$

K představuje matici kamery a R je rotace nalezená RANSAC algoritmem. Na základě tohoto vztahu budou zavedeny vektory $\vec{r}_k = [r_{k1}, r_{k2}, r_{k3}]^T$ a $\vec{s} = [s_1, s_2, s_3]^T$.

$$\vec{r}_k = K \cdot R \cdot \vec{X}_k, \quad \vec{s} = K \cdot \vec{t} \quad (4.8)$$

Soustava rovnic (4.7) může být po dosažení nových vektorů upravena do tvaru:

$$\begin{aligned} \lambda \cdot u_k &= r_{k1} + s_1 \\ \lambda \cdot v_k &= r_{k2} + s_2 \\ \lambda &= r_{k3} + s_3 \end{aligned} \quad (4.9)$$

Do prvních dvou rovnic bude dosazeno za neznámou λ , která je vyjádřena ve třetí rovnici:

$$\begin{aligned} r_{k3} \cdot u_k + s_3 \cdot u_k &= r_{k1} + s_1 \\ r_{k3} \cdot v_k + s_3 \cdot v_k &= r_{k2} + s_2 \end{aligned} \quad (4.10)$$

Pro další výpočty bude tato soustava upravena do tvaru:

$$\begin{aligned} -s_1 + s_3 \cdot u_k &= r_{k1} - r_{k3} \cdot u_k \\ -s_2 + s_3 \cdot v_k &= r_{k2} - r_{k3} \cdot v_k \end{aligned} \quad (4.11)$$

Na základě soustavy rovnic (4.11) budou pro každý z bodů určeny matice A_k a B_k :

$$A_k = \begin{bmatrix} -1 & 0 & u_k \\ 0 & -1 & v_k \end{bmatrix}, \quad B_k = \begin{bmatrix} r_{k1} - r_{k3} u_k \\ r_{k2} - r_{k3} v_k \end{bmatrix} \quad (4.12)$$

Z matic pro jednotlivé body bude sestrojena výsledná přeurčená soustava lineárních rovnic s neznámou \vec{s} ve tvaru:

$$A \cdot \vec{s} = B \quad (4.13)$$

kde matice A a B jsou sestrojeny podle předpisu:

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} \quad (4.14)$$

Soustavu rovnic pro neznámou \vec{s} lze řešit standardní metodou pro řešení soustavy lineárních přeúčených rovnic:

$$\vec{s} = (A^T A)^{-1} A^T B \quad (4.15)$$

Výsledná translace bude vypočtena z vektoru \vec{s} na základě rovnice (4.8):

$$\vec{t} = K^{-1} \cdot \vec{s} \quad (4.16)$$

4.2.5 Určení úhlů z rotační matice

Na závěr lze matici rotace převést na rotační úhly. Přestože vyjádření rotace v úhlech není pro získání transformací mezi snímky nezbytné, umožňuje snazší kontrolu získaných výsledků. Dále toto vyjádření parametrizuje rotaci pouze třemi parametry. Pro vyjádření rotace byla zvolena v robotice standardně používaná soustava úhlů yaw, pitch, roll [37]. Tyto úhly budou dále značeny α , β , γ a rotační matici R lze z těchto úhlů vyjádřit maticí:

$$\begin{bmatrix} \cos(\alpha) \cos(\beta) & \cos(\alpha) \sin(\beta) \sin(\gamma) - \sin(\alpha) \cos(\gamma) & \cos(\alpha) \sin(\beta) \cos(\gamma) + \sin(\alpha) \sin(\gamma) \\ \sin(\alpha) \cos(\beta) & \sin(\alpha) \sin(\beta) \sin(\gamma) + \cos(\alpha) \cos(\gamma) & \sin(\alpha) \sin(\beta) \cos(\gamma) - \cos(\alpha) \sin(\gamma) \\ -\sin(\beta) & \cos(\beta) \sin(\gamma) & \cos(\beta) \cos(\gamma) \end{bmatrix} \quad (4.17)$$

Při opačné úloze, kdy jsou z rotační matice získávány hodnoty jednotlivých úhlů, je pro obecné řešení nutné nejprve zkontrolovat, jestli prvek rotační matice r_{31} (první index představuje řádek a druhý sloupec matice R , ve kterém se daný prvek vyskytuje) není roven 1 nebo -1. Pokud není, lze určit úhly z matice R dále uvedených vztahů (4.18). Opačná situace nebude uvažována, protože by znamenala extrémní překlopení podvozku, které lze v dané aplikaci vyloučit.

$$\begin{aligned} \alpha &= \text{atan2}(r_{21}, r_{11}) \\ \beta &= \arcsin(-r_{31}) \\ \gamma &= \text{atan2}(r_{32}, r_{33}) \end{aligned} \quad (4.18)$$

4. Navržené metody

5 Implementace a experimenty

5.1 Hardware

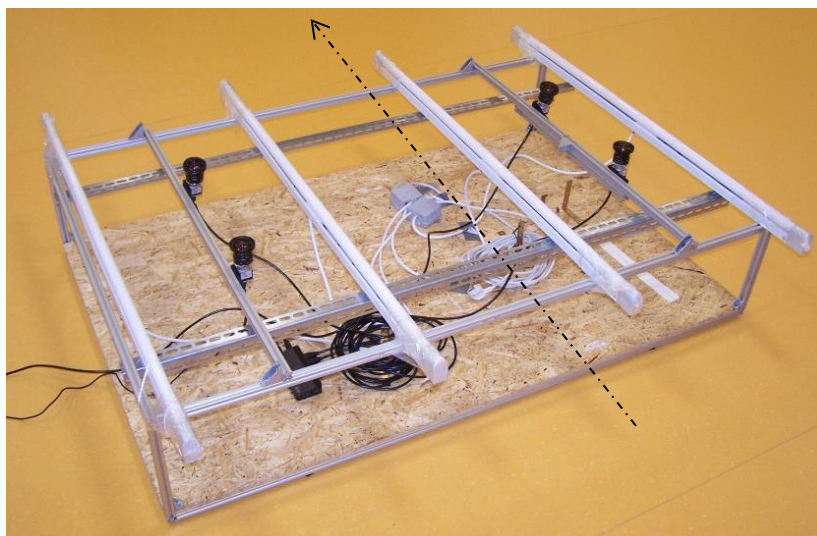
5.1.1 Výpočetní jednotka

Výpočetní jednotkou pro provádění experimentů byl notebook Asus N73SV vybavený procesorem Intel® Core™ i7-2630QM [38]. Procesor disponuje čtyřmi fyzickými jádry, na kterých může být paralelně spuštěno až 8 vláken. Základní takt procesoru je 2GHz. Při zatížení pouze jednoho jádra může dojít k přetaktování až na frekvenci 2,9GHz. Notebook dále disponuje 8GB pamětí RAM. Software byl vyvíjen na operačním systému Ubuntu 16.04 s verzí knihovny OpenCV 3.4.

5.1.2 Kamera Basler acA1920-155uc

Používanými kamerami pro prototyp skeneru podvozků v projektu Cassandra jsou kamery Basler acA1920-155uc [39]. Tato kamera má CMOS snímač s rozlišením 1920x1200px. Důležitou vlastností pro budoucí aplikaci je závěrka typu Global Shutter, která zajišťuje vyčítání všech pixelů ve stejný okamžik. U většiny běžných kamer se používá typ závěrky Rolling Shutter, kdy jsou jednotlivé pixely vyčítány postupně, což při snímání pohybujícího se objektu vede k deformaci obrazu. Další důležitou vlastností, která nebyla vzhledem k statickému snímání testovacích dat v této práci využita, je vysoká snímkovací frekvence dosahující až 164Hz. Kamery byly dále osazeny objektivy se zorným úhlem 90°.

Pro pořízení testovacích dat podvozku byla v rámci projektu sestrojena posuvná konstrukce, na kterou byly uchyceny kamery a osvětlení. Konfigurace kamer byla zvolena tak, aby dva stereo páry kamer pokryly celou šíři podvozku. Jednotlivé kamery stereo páru byly umístěny ve směru jízdy automobilu ve vzdálenosti okolo 30cm od sebe. Aby se snímky více překrývaly, byla jedna z kamer sklopena o 15°. Druhá kamera byla orientována kolmo k podvozku. Stereo páry byly od sebe vzdáleny v příčném směru ke směru pohybu o 1m.



Obr. 9 Konstrukce s kamerami Basler pro pořízení testovacích snímků podvozku automobilu. Šipka znázorňuje směr pohybu konstrukce při snímání podvozku automobilu.

5.1.3 Dataset pro testování

Pro provádění testů byla pořízena sada snímků podvozku za použití konstrukce s kamerami představené v předchozí kapitole. Konstrukce byla posouvána pod zvednutým automobilem v podélném směru s posunem 5cm mezi po sobě jdoucími snímky. Tyto vzdálenosti byly určeny s přesností pod 2mm. V testovacích datech byly zachyceny dvě různé výšky podvozku ve vzdálenostech okolo 57 a 75cm od kamery. Aby byly simulovány i různé světelné podmínky, byly zároveň snímky pořízeny se dvěma různými časy expozice (3ms a 10ms). Vzhledem ke krátké délce kolejnic pro posun konstrukce s kamerami, byla pro každou výšku podvozku i dobu expozice nasnímána zvlášť přední i zadní část automobilu. Na závěr bylo s kratším časem expozice pořízeno několik snímků s různým úhlem natočení (úhel yaw). Měření úhlu probíhalo řízeným vychýlením konstrukce z jedné kolejnice. Výsledný úhel byl určen na základě trojúhelníku tvořeného kolejnicemi a konstrukcí s kamerami. Odchylna referenčního měření úhlu byla pod 2°. Snímky s kratší dobou expozice budou dále označovány jako tmavé a snímky s delším časem světlé.

5.2 Hledání optimální metody pro detekci a popis příznaků

5.2.1 Automatické testování a testovací množina

Optimální metoda pro detekci a popis příznaků byla hledána automatickým testem parametrů, který je podrobněji popsán v kapitole 4.1. Jednotlivé hodnoty parametrů byly zvoleny na základě experimentálního testování na vybraných snímcích z testovací množiny. Počet hodnot pro jeden parametr byl volen s ohledem na celkovou náročnost výpočtů tak, aby celková doba výpočtu pro všechny kombinace zvolené metody nebyla delší než několik hodin.

Data pro testování byla vybírána z pořízené sady snímků podvozku automobilu tak, aby obsahovala všechny části podvozku, pro které může být charakter snímané scény velmi odlišný. Nižší velikosti testovací množiny bylo dále dosaženo vybráním snímků pouze z jednoho stereopáru kamer. Pro každou kombinaci času expozice a vzdálenosti bylo do výsledné testovací sady vybráno 5 dvojic snímků. Kromě celkového výsledku testování byly jednotlivé typy snímků ohodnoceny zvlášť. Ve výsledcích jsou uváděny dohromady výsledky pro světlé a tmavé snímky. Rozdělování na základě vzdálenosti nebylo rozlišováno, protože v reálném nasazení systém bude muset pracovat s různými výškami podvozků aut.

Při výběru optimálních parametrů byla největší váha přikládána minimálnímu počet správně přiřazených příznaků v jednom snímku. Dále v pořadí byl uvažován maximální čas detekce a přiřazení příznaků a celkový počet správných přiřazení. Vzhledem k požadavku na zpracování v čase blízkém reálnému, byla preferována řešení s maximálním časem výpočtu pod 1s. Aby bylo možné hodnoty pro tmavé a světlé snímky porovnávat s celkovým výsledkem, jsou počet správně přiřazených příznaků a počet detekovaných příznaků ve všech snímcích pro tmavé a světlé snímky, které jsou součtem pouze deseti hodnot, vynásobeny dvěma.

5.2.2 Harrisův detektor a BRIEF deskriptor

První testovanou metodou pro extrakci příznaků ze snímku byl Harrisův detektor (kapitola 3.4.1) s BRIEF deskriptorem (kapitola 3.4.6). V knihovně OpenCV je Harrisův detektor rohů implementován ve funkci *cornerHarris* nebo *goodFeaturesToTrack*, která je používána v této práci. Tato funkce má několik volitelných parametrů. Nastavit lze parametr *k* a různou

velikost okna. Pro výpočet diferencí se používá Sobelův filtr, u kterého lze zvolit velikost okna pro výpočet diference. Podporované hodnoty jsou pouze 1, 3, 5 a 7. Dále lze určit minimální vzdálenost mezi rohy, čímž by se mělo zabránit existenci více příznaků pro sejný roh. Nastavuje se také maximální počet rohů, které má funkce vrátet. Relativně vůči tomuto počtu je možné zvolit minimální úroveň kvality vrácených příznaků.

Výpočet BRIEF deskriptoru je implementován v rozšiřující knihovně OpenCV *xfeatures2D* ve třídě *BriefDescriptorExtractor*. V nastavení lze zvolit velikost deskriptoru 16, 32 nebo 64 byte a nastavit, jestli mají být deskriptory invariantní vůči rotaci. Rotace může být kompenzována z vlastností nalezených bodů před samotným výpočtem deskriptoru, pokud je informace o rotaci získávána při detekci příznaku.

Testovanými parametry byly minimální vzdálenost mezi příznaky, relativní kvalita příznaku, velikost okna, parametr k a velikost okna pro Sobelův filtr. Nastavení zbylých parametrů bylo zvoleno přímo. Velikost deskriptoru byla 64byte, aby byla dimenze prohledávaného prostoru nižší. Použití orientace při výpočtu deskriptoru bylo zakázáno, protože v Harrisově detektoru není počítána. Maximální počet příznaků byl nastavován pro všechny metody, u kterých bylo možné zvolit počet příznaků, shodně na 300 příznaků v jedné části snímku. V celém snímku mohlo být detekováno maximálně 3600 příznaků. Hodnoty všech parametrů pro testování jsou uvedeny v Tab. 1. Celkový počet testovaných kombinací byl 1500.

Parametry pro Harrisův a Shi-Tomasi detektor a BRIEF deskriptor	
Parametr	Testované hodnoty
Minimální vzdálenost	3, 5, 10
Relativní kvalita příznaku	0.1, 0.01, 0.001, 0.0001, 0.00001
Velikost okna	11, 21, 31, 41, 51
Okno Sobelův filtr	1, 3, 5, 7
Parametr k (pouze Harris)	0.01, 0.005, 0.001, 0.0005, 0.0001
Počet příznaků	300
Velikost deskriptoru	64
Kompenzace rotace	NE

Tab. 1 Testované parametry pro Harrisův a Shi-Tomasi detektor a BRIEF deskriptor.

Nalezené optimální hodnoty parametrů z testovaných kombinací jsou uvedeny v Tab. 2. Ze získaných výsledků vyplývá, že Harrisův detektor není vhodnou metodou, protože nezaručuje dostatečný počet bodů pro přiřazování mezi snímky stereo páru. Z velkého rozdílu celkového počtu přiřazení a počtu správně přiřazených příznaků vyplývá, že počet detekovaných příznaků je dostatečný, ale získané příznaky jsou obtížně rozlišitelné.

Optimální hodnoty parametrů pro Harrisův detektor a BRIEF deskriptor			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	5	6	4
Správná přiřazení celkem	162	176	150
Přiřazení celkem	21 492	25 512	23 787
Max. čas detekce a přiřazení [s]	0,26	0,26	0,28
Minimální vzdálenost	3	5	3
Relativní kvalita příznaku	0,0001	0,00001	0,00001
Velikost okna	11	11	21
Parametr k	0,005	0,0005	0,0001
Okno Sobelův filtr	5	5	5

Tab. 2 Optimální hodnoty parametrů a výsledky automatického testu pro Harrisův detektor a BRIEF deskriptor

5.2.3 Shi-Tomasi detektor a BRIEF deskriptor

Dalším testovaným detektorem byl Shi-Tomasi detektor (kapitola 3.4.2), jehož příznaky byly popsány BRIEF deskriptory. V knihovně OpenCV je Shi-Tomasi detektor implementován stejně jako Harrisův detektor ve funkci *goodFeaturesToTrack*. Možnosti nastavení jsou proto stejné až na parametr k , který je určen pouze pro Harrisův detektor.

Také hodnoty testovaných parametrů byly stejné jako u Harrisova detektoru v předchozí kapitole až na parametr k , který není používán. Seznam testovaných hodnot parametrů je uveden v Tab. 1. Celkový počet testovaných kombinací byl 300.

Nejlepší nalezené parametry jsou uvedeny v tabulce Tab. 3. Ani Shi-Tomasi detektor nedokáže nalézt dostatečný počet kvalitních příznaků. Z porovnání počtu správně přiřazených příznaků k celkovému počtu vyplývá, že pouze každé 154 přiřazení lze označit za správné a nalezené příznaky jsou příliš slabé.

Optimální hodnoty parametrů pro Shi-Tomasi detektor a BRIEF deskriptor			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	4	5	4
Správná přiřazení celkem	154	190	161
Přiřazení celkem	17 146	24 746	24 892
Max. čas detekce a přiřazení [s]	0,26	0,29	0,29
Minimální vzdálenost	3	3	3
Relativní kvalita příznaku	0,1	0,0001	0,0001
Velikost okna	31	31	31
Okno Sobelův filtr	3	3	3

Tab. 3 Optimální hodnoty parametrů a výsledky automatického testu pro Shi-Tomasi detektor a BRIEF deskriptor

5.2.4 SIFT

Metoda SIFT (kapitola 3.4.3) je implementována v rozšíření *xfeatures2D* knihovny OpenCV. Nastavitelnými parametry metody jsou počet vrstev v jednotlivých oktávách, hranice pro zamítnutí slabých příznaků, parametr r pro odstranění hran, počáteční hodnota parametru σ a počet nejlepších příznaků, který má metoda vrátit. Počet oktáv je vybírán automaticky na základě rozlišení vstupních snímků.

Testovanými parametry byly všechny zmíněné s výjimkou maximálního počtu vrácených příznaků, který byl nastaven na hodnotu 300. Přehled testovaných hodnot je uveden v Tab. 4. Testováno bylo 288 kombinací parametrů.

Parametry SIFT	
Parametr	Testované hodnoty
Počet vrstev oktávy	1, 2, 3
Hranice síly příznaku	0.05, 0.03, 0.01, 0.008, 0.005, 0.001
Parametr r	20, 50, 100, 200
Počáteční σ	1.2, 1.4, 1.6, 1.8
Počet příznaků	300

Tab. 4 Testované parametry pro metodu SIFT

Nejlepší parametry nalezené při automatickém testování jsou uvedeny v Tab. 5. Nejvýraznějším nedostatkem metody SIFT byla maximální doba detekce a přiřazení příznaků, která pro žádný z testovaných stereo snímků nebyla nižší než 1,2s. Nejlepší nalezené parametry pro světlé snímky dokonce dosáhly maximálního času výpočtu téměř 5s. Vzhledem

k dlouhému času výpočtů není možné tuto metodu využít pro metodu vizuální odometrie navrhované v této práci, přestože dokáže v každém stereo snímku nalézt dostatečný počet správně přiřazených příznaků.

Optimální hodnoty parametrů pro metodu SIFT			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	137	258	137
Správná přiřazení celkem	3 748	6 162	4 737
Přiřazení celkem	35 718	36 276	35 701
Max. čas detekce a přiřazení [s]	2,13	4,97	2,13
Počet vrstev oktávy	2	1	2
Hranice síly příznaku	0,005	0,001	0,005
Parametr r	20	20	20
Počáteční σ	1,8	1,8	1,8

Tab. 5 Optimální hodnoty parametrů a výsledky automatického testu pro metodu SIFT

5.2.5 SURF

Také metoda SURF (kapitola 3.4.4) je implementována v rozšíření *xfeatures2D* knihovny OpenCV. Jejími volitelnými parametry jsou počet oktáv, počet vrstev v jedné oktávě a minimální úroveň pro determinant Hessovy matice, podle které jsou slabé příznaky filtrovány. Dále lze zvolit, jestli má být používán deskriptor s 64 nebo 128 prvky a počítána orientace příznaků (použití U-SURF). Maximální počet vrácených příznaků u této metody zvolit nelze.

Testovány byly všechny volitelné parametry včetně obou binárních pro volbu délky deskriptoru a použití metody U-SURF. Dohromady bylo testováno 384 kombinací. Hodnoty parametrů jsou uvedeny v Tab. 6.

Parametry SURF	
Parametr	Testované hodnoty
U-SURF	ANO/NE
Rozšířený deskriptor	ANO/NE
Počet oktáv	1, 2, 3, 5
Počet vrstev oktávy	1, 2, 3, 5
Min. determinant	20, 50, 80, 100, 150, 200

Tab. 6 Testované parametry pro metodu SURF

Optimální hodnoty parametrů pro metodu SURF			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	49	356	49
Správná přiřazení celkem	2 384	9 822	2 384
Přiřazení celkem	24 310	57 834	24 310
Max. čas detekce a přiřazení [s]	0,49	0,96	0,96
U-SURF	ANO	ANO	ANO
Rozšířený deskriptor	NE	NE	NE
Počet oktáv	1	1	1
Počet vrstev oktávy	1	1	1
Min. determinant	20	20	20

Tab. 7 Optimální hodnoty parametrů a výsledky automatického testu pro metodu SURF

5. Implementace a experimenty

Při vyhodnocení výsledků metody SURF nebyly uvažovány parametry, pro které maximální čas výpočtu překročil 1s. Celkově nejlepší minimální počet přiřazených příznaků byl 191, pro který ale maximální čas detekce trval 9,24s. Metoda s nejlepšími nalezenými parametry, které jsou uvedeny v Tab. 7, detekuje dostatečný počet příznaků v zadaném limitu 1s pro maximální čas detekce a přiřazení. Proto je jednou z možných variant pro navrhovanou metodu vizuální odometrie.

5.2.6 FAST detektor a BRIEF deskriptor

Dalším testovaným detektorem rohů s BRIEF deskriptorem byl FAST detektor (kapitola 3.4.5). Základní varianta FAST detektoru je implementována v knihovně OpenCV ve funkci *FAST*. Nastavitelnými parametry detektoru jsou parametr t , volba použití Non-maximal Suppression a výběr jednoho ze tří možných typů detektoru s danou velikostí kružnice a parametrem n . Možné rozměry jsou 16, 12 a 8 pixelů, ke kterým je stanoveno n na hodnoty 9, 7 a 5. Maximální počet vrácených příznaků nastavit nelze.

Testovány byly všechny volitelné parametry pro FAST detektor a různé velikosti deskriptoru pro BRIEF deskriptor. Volba kompenzace orientace nebyla používána, protože není FAST detektorem počítána. Hodnoty jednotlivých parametrů ze 144 testovaných kombinací jsou uvedeny v Tab. 8.

Parametry pro FAST detektor s BRIEF deskriptorem	
Parametr	Testované hodnoty
Parametr t	3, 5, 10, 15, 18, 20, 25, 30
Non-maximal Suppression	ANO/NE
Typ [px]	16, 12 a 8
Velikost deskriptoru [byte]	16, 32, 64
Kompenzace rotace	NE

Tab. 8 Testované parametry pro FAST detektor s BRIEF deskriptorem

Obdobně jako u SURF detektoru byla vybírána pouze řešení, pro která maximální čas zpracování stereo snímku nepřesáhl 1s. Obdobně jako Harrisův a Shi-Tomasi detektor FAST detektor nalezne dostatečný počet příznaků, které jsou ovšem příliš slabé. Proto není tento detektor vhodný pro metodu navrhovanou v této práci. Nejlepší nalezené parametry jsou uvedeny v Tab. 9.

Optimální hodnoty pro FAST detektor s BRIEF deskriptorem			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	0	9	0
Správná přiřazení celkem	104	590	371
Přiřazení celkem	8 464	52 566	29 950
Max. čas detekce a přiřazení [s]	0,072	0,68	0,77
Parametr t	15	15	15
Non-maximal Suppression	ANO	ANO	ANO
Typ [px]	12	12	12
Velikost deskriptoru [byte]	32	16	32

Tab. 9 Optimální hodnoty parametrů a výsledky automatického testu pro FAST detektor s BRIEF deskriptorem

5.2.7 ORB

Poslední testovanou metodou je ORB (kapitola 3.4.7), která je v knihovně OpenCV implementována ve stejnojmenné třídě *ORB*. Detekce postavená na FAST detektoru má pevně nastavený obvod kružnice 16px a parametr $n=9$ (viz kapitola 3.4.5). Pro hledání a popis příznaků je možné nastavit celou řadu parametrů. Základními parametry jsou maximální počet nalezených příznaků, počet jednotlivých vrstev a proporcionální změna rozměrů snímku mezi jednotlivými vrstvami. Nastavit lze i parametr t pro FAST detektor. Pro BRIEF deskriptor je možné zvolit velikost oblasti, ve které budou hledány body binárního testu.

OpenCV rovněž umožňuje použití 3 nebo 4 bodů pro výpočet jednotlivých částí BRIEF deskriptoru. Při této změně však už není deskriptor binární a musí se pro výpočet vzdálenosti použít speciální varianta Hammingovy vzdálenosti. I v tomto případě je však prvek deskriptoru určen maximem jedno z trojice nebo čtveřice bodů. Pro získání příznaků blíže středu snímku lze nastavit hranici od okraje, ve které nebudou příznaky detekovány. Posledním parametrem je výběr metody pro ohodnocení příznaků. Volit lze mezi ohodnocením z Harrisova detektoru nebo ohodnocením podle funkce V , která se používá u FAST detektoru.

Testovanými parametry pro metodu ORB byly počet vrstev, změna rozměrů mezi vrstvami, parametr t , velikost oblasti pro výpočet BRIEF deskriptoru a typ ohodnocení z Harrisova nebo FAST detektoru. Zbylé parametry byly nastaveny fixně. Všechny testované hodnoty parametrů jsou uvedeny v Tab. 10.

Parametry ORB	
Parametr	Testované hodnoty
Počet vrstev	1, 3, 5, 8, 10
Změna mezi vrstvami	1.2, 1.4, 1.6, 1.8
Velikost oblasti pro BRIEF [px]	15, 21, 31, 41
Ohodnocení	Harris/FAST
Parametr t	5, 8, 10, 15, 20
Počet příznaků	300
Body pro 1 prvek deskriptoru	2
Vzdálenost od okraje snímku	0

Tab. 10 Testované parametry pro metodu ORB

Výsledky automatického testu jsou uvedeny v Tab. 11. Metoda dokáže najít a správně přiřadit dostatečný počet příznaků ve všech snímcích nezávisle na osvětlení. Zároveň jsou detekce i přiřazení příznaků dostatečně rychlé, a proto je tato metoda vhodná pro vizuální odometrii aplikovanou na podvozky automobilů.

Optimální hodnoty pro metodu ORB			
	Tmavé snímky	Světlé snímky	Celkem
Minimální počet příznaků	158	194	158
Správná přiřazení celkem	4 402	5 198	4 899
Přiřazení celkem	41 756	37 570	39 773
Max. čas detekce a přiřazení [s]	0,38	0,36	0,39
Počet vrstev	5	8	5
Změna mezi vrstvami	1,2	1,2	1,2
Velikost oblasti pro BRIEF [px]	31	31	31
Ohodnocení	FAST	FAST	FAST
Parametr t	5	8	5

Tab. 11 Optimální hodnoty parametrů a výsledky automatického testu pro ORB metodu

5.2.8 Porovnání testovaných metod

Z testovaných metod pro detekci a popis příznaků není možné většinu metod použít. Hlavními důvody jsou nedostatečný počet správných přiřazení mezi snímky a příliš dlouhá doba pro detekci příznaků, vytváření deskriptorů a přiřazování příznaků mezi snímky. Jedinými metodami v automatickém testování, které splňovaly zadané parametry, byly metody SURF a ORB. ORB je oproti metodě SURF rychlejší a zároveň dosahuje vyšší hodnoty minimálního počtu příznaků pro jeden snímek. Metoda SURF má naopak tento parametr vyšší pro světlé snímky. Pro vizuální odometrii bylo zvoleno použití ORB deskriptoru zejména kvůli nižší časové náročnosti a větší robustnosti pro tmavé snímky. Zvolené parametry odpovídají nejlepším společným parametrům uvedeným v Tab. 11 pro světlé i tmavé snímky.

5.3 Ověření navržené metody na reálných datech

5.3.1 Parametry výsledné metody

Na základě teoretické části popsané v kapitole 3 a automatického testu příznaků, který je popsán v předchozí kapitole, byla navržena metoda pro vizuální odometrii (kapitola 4.2). Pro tuto metodu bylo potřeba určit několik různých parametrů tak, aby metoda pracovala co nejrobustněji pro všechny typy snímků a velikosti posunů mezi nimi. Určování těchto parametrů probíhalo zejména experimentálním testováním, při kterém byl důraz kladen nejen na zpracování všech snímků, ale i na časovou náročnost.

V nastavení parametrů metody ORB byla upravena pouze hodnota maximálního počtu vrácených příznaků, která byla zvýšena na 500 oproti 300 příznakům používaným v automatickém testu. Tento počet příznaků je vztažen k jednotlivým částem snímku, kterých bylo celkem 12. Výsledný počet příznaků v jednom snímku může být až 6000. Hodnoty parametrů metody jsou zaznamenány v následující tabulce:

Parametry navržené metody pro vizuální odometrii	
Parametr	Hodnota
Maximální počet detekovaných příznaků pro 1 oblast	500
Rozdělení snímku pro detekci příznaků	4x3
Maximální vzdálenost příznaku od epipoláry	5px
Hranice pro filtraci na základě vzdálenosti	5px
Počet RANSAC iterací celkem	700
Počet vláken pro RANSAC	4
Hranice pro reprojekční chybu u RANSAC algoritmu	5px

Tab. 12 Parametry navržené metody pro vizuální odometrii popsané v kapitole 4.2

5.3.2 Výsledky testování pro světlé snímky

Navržená metoda byla na rozdíl od automatického testu testována na datech z obou párů kamer. Celkem byly testovány 3 různé vzdálenosti mezi snímky - 10, 20 a 30 cm. Vyhodnocovanými parametry byly průměrná chyba translace mezi snímky, směrodatná odchylka, maximální chyba translace a maximální čas výpočtu posunu mezi dvěma snímky. Dále byla maximální chyba vyjádřena procentuálně vůči skutečnému posunu. Výsledky metody testované na testovacím datasetu pro světlé snímky jsou uvedeny v Tab. 13

Z naměřených hodnot vyplývá, že chyba určení posunu pro světlé snímky nepřekročí 3%. Maximální průměrná chyba je 0,16cm. Pokud by tato hodnota platila pro celý 4,5m dlouhý podvozek složený z 20 snímků, byla by výsledná chyba v délce podvozku 3,04cm, což je 0,7%.

Při návrhu metody byl větší důraz kladen na přesnost pro světlé snímky, protože ve výsledném systému pro projekt KASSANDRA bude dostupné umělé osvětlení, které by mělo zajistit dostatečnou světelnost pořízených snímků a eliminovat výrazné výkyvy intenzity.

Maximální čas výpočtu 1,36s by při dvaceti snímcích odpovídal času výpočtu 25,84s pro celý podvozek. U výsledné metody je možné počítat se snížením tohoto času, protože nebude potřeba znovu počítat všechny příznaky v každé iteraci. Podrobněji je tato optimalizace rozebrána v kapitole 5.4.

Výsledky testování navržené metody pro světlé snímky							
Stereo pár	Vzdálenost podvozku [cm]	Posun mezi snímky [cm]	Průměrná chyba [cm]	Směrodatná odchylka [cm]	Maximální chyba [cm]	Maximální chyba vůči posunu [%]	Maximální čas výpočtu [s]
1	57	10	0,03	0,10	0,29	2,9	1,261
1	75	10	0,04	0,13	0,23	2,3	1,153
1	57	20	0,06	0,16	0,45	2,2	1,239
1	75	20	0,08	0,22	0,35	1,7	1,111
1	57	30	0,09	0,22	0,52	1,7	1,161
1	75	30	0,14	0,30	0,45	1,5	1,043
2	57	10	0,03	0,09	0,18	1,8	1,264
2	75	10	0,05	0,08	0,14	1,4	1,360
2	57	20	0,08	0,14	0,25	1,2	1,214
2	75	20	0,10	0,13	0,25	1,2	1,264
2	57	30	0,13	0,19	0,35	1,2	1,128
2	75	30	0,16	0,18	0,30	1,0	1,200
Celkem:			0,08	0,17	0,52	2,9	1,360

Tab. 13 Výsledky experimentálního ověření navržené metody pro světlé snímky

Výsledky testování navržené metody pro tmavé snímky							
Stereo pár	Vzdálenost podvozku [cm]	Posun mezi snímky [cm]	Průměrná chyba [cm]	Směrodatná odchylka [cm]	Maximální chyba [cm]	Maximální chyba vůči posunu [%]	Maximální čas výpočtu [s]
1	57	10	0,02	0,15	0,67	6,7	1,130
1	75	10	0,03	0,10	0,21	2,1	0,998
1	57	20	0,05	0,20	0,72	3,6	1,098
1	75	20	0,05	0,17	0,38	1,9	0,937
1	57	30	0,10	0,25	0,78	2,6	1,032
1	75	30	0,09	0,24	0,45	1,5	0,914
2	57	10	0,05	0,13	0,47	4,7	1,150
2	75	10	0,05	0,08	0,24	2,4	1,185
2	57	20	0,12	0,19	0,40	2,0	1,081
2	75	20	0,10	0,14	0,33	1,6	1,131
2	57	30	0,20	0,24	0,41	1,4	1,026
2	75	30	0,16	0,19	0,36	1,2	1,040
Celkem:			0,09	0,18	0,78	6,7	1,185

Tab. 14 Výsledky experimentálního ověření navržené metody pro tmavé snímky

5.3.3 Výsledky testování pro tmavé snímky

Tmavé snímky byly testovány obdobně jako světlé snímky v předchozí kapitole. Výsledky testu pro tmavé snímky jsou uvedeny v Tab. 14.

Přestože byla metoda navrhována převážně pro světlé snímky, určila translaci pro všechny tmavé snímky s průměrnou přesností pod 1mm. Oproti světlým snímkům byly zaznamenány výraznější odchylky od referenčních vzdáleností s maximální chybou určení translace 6,7%.

5.3.4 Rotace

Pro testování rotace byla pořízena sada několika tmavých snímků, ve kterých je simulována rotace podél osy z (úhel yaw) spojená s mírným posunem v podélném směru. Metoda dokázala bezchybně určit velikost rotace pro všechny úhly do maximálního měřeného rozsahu 15°. Maximální odchylka od referenční hodnoty úhlu byla pod 1,5°. Tyto naměřené přesnosti jsou dokonce nižší než přesnost, se kterou byla určována velikost referenčních úhlů. Celkové výsledky testu rotace jsou uvedeny v Tab. 15.

Výsledky navržené metody pro rotaci v úhlu yaw			
Stereo pár	Rotace [°]	Naměřená rotace [°]	Chyba rotace [°]
1	3,2	3,1	0,1
2	3,2	3,0	0,2
1	6,3	5,9	0,4
2	6,3	5,6	0,7
1	9,5	9,1	0,4
2	9,5	9,1	0,4
1	12,5	11,1	1,4
2	12,5	11,0	1,5
1	15,5	14,3	1,2
2	15,5	14,0	1,5

Tab. 15 Testování přesnosti určení rotace

5.4 Analýza výpočetní náročnosti navržené metody

Na závěr bylo provedeno měření výpočetní náročnosti jednotlivých kroků navržené metody pro vizuální odometrii. Tyto hodnoty jsou uvedeny v Tab. 16. Měření bylo prováděno pro světlé snímky se vzdáleností podvozku od kamery 75cm. Nejvíce výpočetně náročnou operací byl RANSAC algoritmus, který se podařilo výrazně zrychlit rozdělením do více vláken. Prostor pro další optimalizaci je zejména u přiřazování a detekce příznaků.

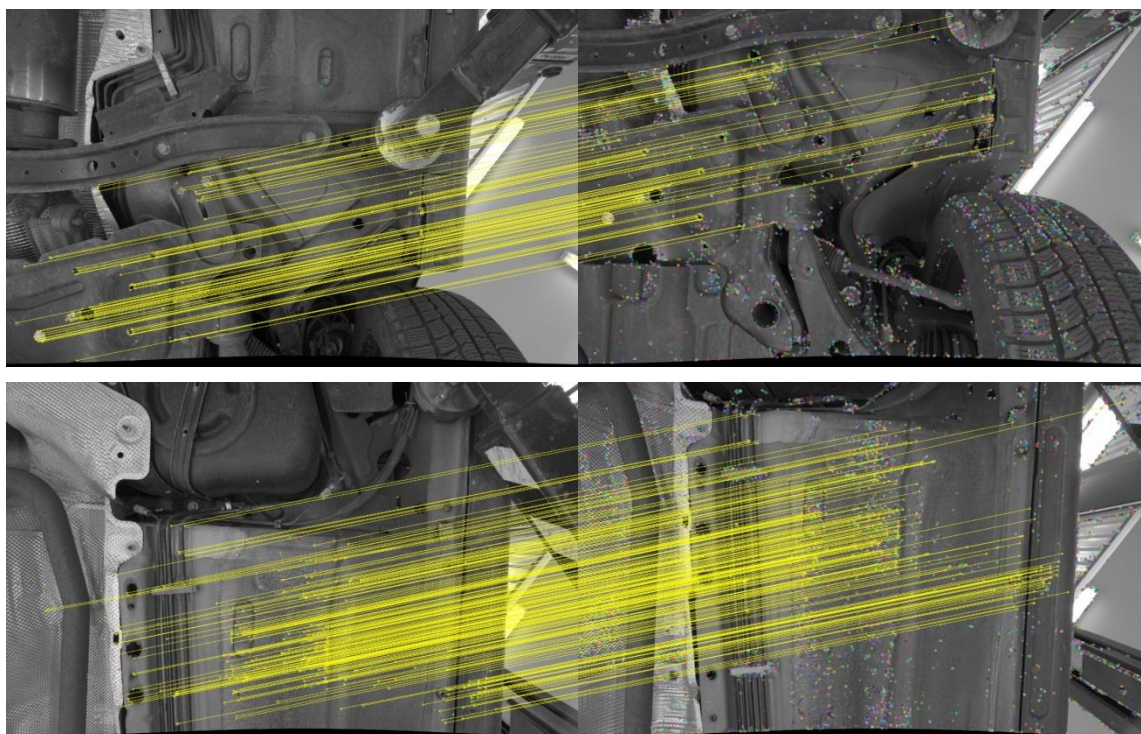
Analýza výpočetní náročnosti navržené metody	
Krok metody	Čas [s]
Detekce a popis příznaků ve snímcích 1, 2, 3	0,425
Přiřazování příznaků mezi snímky 1-2	0,238
Filtrování podle vzdálenosti od epipoláry a triangulace	0,035
Přiřazování mezi snímky 1-3	0,018
Triangular matching (+ přiřazení 2-3)	0,020
Filtrování příznaků podle vzdálenosti	0,001
RANSAC (4 vlákna)	0,243
Výpočet translace	0,000
Celkem	0,984

Tab. 16 Průměrný čas výpočtu jednotlivých kroků metody pro světlé snímky se vzdáleností podvozku 75cm

Navržená metoda byla zatím implementována jen pro výpočet posunu mezi dvěma snímky. Při kontinuálním snímání celého podvozku bude možné provést několik optimalizací, které výrazně sníží výpočetní náročnost metody. Při kontinuálním snímání lze provádět detekci a přiřazování příznaků pro oba stereo snímky pouze v každém lichém kroku. Každé druhé přiřazení potom bude prováděno v opačném směru, kdy budou triangulovány body ze snímků v čase i a promítány zpět do snímku v čase $i-1$. K takto získané transformaci bude pouze spočítána inverzní transformace odpovídající výsledné transformaci mezi snímky. Tímto přístupem bude možné snížit počet přiřazování mezi snímky stereo páru pro podvozek skládající se z 20 snímků o 9. Při průměrném času výpočtu přiřazení s triangulací 0,273s jde o úsporu 2,45s.

Pokud budou zároveň detekované příznaky mezi jednotlivými kroky předávány, je pro dvacet snímků potřeba detekovat příznaky pouze 30x, oproti 57 detekcím v aktuální implementaci. Pokud detekce a popis příznaků v jednom snímku trvají průměrně 0,142s, jde o úsporu 3,834s

Pro naměřený průměrný čas 0,984s by bez optimalizací určení posunu pro celý podvozek trvalo 18,7s. Použitím jednoduchých optimalizací popsaných v této části by bylo možné snížit čas výpočtu na 12,4s, což odpovídá zrychlení o 33,7%.



Obr. 10 Ukázka snímků podvozku s přiřazenými příznaky mezi 2 světlými posunutými snímky se vzdáleností podvozku 57cm, zobrazeny jsou příznaky se správným přiřazením podle RANSAC algoritmu, ze kterých bude dále určována translace.

5. Implementace a experimenty

6 Závěr

Cílem práce bylo navrhnout a ověřit metodu vizuální odometrie pro dynamickou rekonstrukci 3D modelu podvozků v projektu KASSANDRA. Data získaná z odometrie budou používána ke spojování dílčích 3D modelů podvozku vytvořených ze stereo snímků.

První část práce (kapitola 3) je věnována teoretickému úvodu do vizuální odometrie založené na příznacích a analýze vlastností jednotlivých metod pro vizuální odometrii. Dále jsou v této části popisovány metody pro detekci, popis a přiřazování příznaků a vybrané existující systémy pro vizuální odometrii.

Na základě postupů a metod popsanych v kapitole 3 byla navržena metoda pro vizuální odometrii (kapitola 4.2), která je založena na silné filtraci příznaků podobně jako metoda popsaná v článku [29]. Tento přístup je vhodný i pro příznakově chudé scény, protože pro určení posunu vyžaduje minimální počet příznaků. Pokud budou všechna přiřazení správná, stačí pro určení pozice navrženou metodou 4 body.

Kromě již zmíněné silné filtrace příznaků využívá navržená metoda odděleného výpočtu rotace a translace. Pro výpočet rotace je používán RANSAC algoritmus, ve kterém je rotace určována řešením problému tří bodů. Translace je počítána zvlášť z bodů, které jsou RANSAC algoritmem označeny za správné. Pro detekci a popis příznaků byla zvolena metoda ORB na základě výsledků automatického testu příznaků (kapitola 4.1 a 5.2), kterým byly testovány jednotlivé metody pro detekci a popis příznaků a jejich parametry. Při vyhodnocení testu byl zohledňován požadavek na rychlost metody.

Navržená metoda byla otestována na pořízeném datasetu snímků podvozku automobilu (kapitola 5.3). Tento dataset obsahoval snímky s dvěma úrovněmi jasu. Vzhledem k předpokládanému stabilnímu osvětlení snímané scény pro senzor vyvíjený v projektu KASSANDRA, byla zvolená metoda navrhována primárně pro světlé snímky, u kterých dosahuje maximální chyby translace pod 3%, přičemž průměrná chyba určení polohy je 0,8mm. Na tmavých snímcích je metoda méně stabilní, ale přesto dosahuje průměrné chyby určení posunu pod 1mm s maximální chybou 6,7%.



Obr. 11 Ukázka způsobu snímání podvozku automobilu pro získání experimentálních dat

Kromě tří různých vzdáleností mezi snímky, které představují různé rychlosti automobilu, byla testována i přesnost určení rotace v úhlu yaw. Metoda určila správně rotaci pro všechny testované úhly s maximální rotací $15,5^\circ$. Dosažená přesnost pod $1,5^\circ$ byla pod rozlišovací schopností samotného určení úhlu při provádění experimentu. Na závěr byla s ohledem na budoucí rozvoj metody provedena analýza výpočetní náročnosti jednotlivých kroků navržené metody. V této části byly zároveň analyzovány možnosti zrychlení navržené metody v případě aplikace na souvislou sadu snímků podvozku, u které nebude nutné provádět triangulaci ve všech krocích, a zároveň budou moci být předávány detekované příznaky mezi sousedními kroky. Touto jednoduchou optimalizací bude možné zrychlit navrženou metodu přibližně o 33%.

6.1 Další rozvoj navržené metody

Metoda navržená v této práci je pouze prvním návrhem metody pro prototyp skeneru podvozků v projektu KASSANDRA. Jejím největším nedostatkem je otestování pouze na jednom podvozku, protože obsáhlejší dataset snímků podvozků v době řešení práce nebyl k dispozici. Otestování metody na více podvozcích bude pro její další rozvoj nezbytné. V současném stavu nelze vyloučit přeučení navržené metody na konkrétní typ podvozku. Dále je potřeba metodu více optimalizovat, aby byl snížen její výpočetní čas. Jednou z možností je provádění řady operací místo procesoru na grafické kartě, čímž by mohlo dojít i k řádovému zrychlení některých částí navržené metody. Důležitým krokem také bude spojení vizuální odometrie s vytvářením 3D mapy podvozku do jednoho celku.

Literatura

- [1] KOŠNAR, Karel, Tomáš KRAJNÍK a Libor PŘEUČIL. Kerberos - 3D undercarriage inspection system. 2013. Dostupné také z: <http://imr.ciirc.cvut.cz/uploads/Industry/kerberos.pdf>
- [2] SCARAMUZZA, Davide, Friedrich FRAUNDORFER a Wonpil YU. Visual Odometry [Tutorial]. *International Journal of Control, Automation and Systems*. 2011, 18(4), 80-92. DOI: 10.1109/MRA.2011.943233. ISSN 1070-9932. Dostupné také z: <http://ieeexplore.ieee.org/document/6096039/>
- [3] NISTER, D., O. NARODITSKY a J. BERGEN. Visual odometry. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, 652-659*. DOI: 10.1109/CVPR.2004.1315094. ISBN 0-7695-2158-4. Dostupné také z: <http://ieeexplore.ieee.org/document/1315094/>
- [4] Camera Calibration and 3D Reconstruction. OpenCV 2.4.13.6 documentation [online]. 19. 3. 2018 [cit. 2018-03-20]. Dostupné z: https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html
- [5] PAJDLA, Tomáš. Elements of Geometry for Computer Vision [online]. Publikováno 9. 5. 2016 [cit. 2018-03-20]. Dostupné z: <https://cw.fel.cvut.cz/wiki/lib/exe/fetch.php?tok=d66353&media=http%3A%2F%2Fcmp.felk.cvut.cz%2F~pajdla%2Fgvg%2FGVG-2016-Lecture.pdf>
- [6] HARTLEY, Richard. a Andrew. ZISSERMAN. *Multiple view geometry in computer vision*. 2nd ed. New York: Cambridge University Press, 2003. ISBN 978-0-521-54051-3.
- [7] ŠÁRA, Radim. Anglicko-český a česko-anglický slovníček 3D počítačového vidění [online]. Vydáno 29. prosince 2010 [cit. 2018-03-22]. Dostupné z: <http://cmp.felk.cvut.cz/cmp/courses/TDV/2010W/lectures/3DV-slovník.pdf>
- [8] OpenCV: Open Source Computer Vision Library [online]. [cit. 2018-03-23]. Dostupné z: <https://github.com/opencv/opencv>
- [9] RICHARDSON, Andrew, Johannes STROM a Edwin OLSON. AprilCal: Assisted and repeatable camera calibration. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013, 2013, , 1814-1821*. DOI: 10.1109/IROS.2013.6696595. ISBN 978-1-4673-6358-7. Dostupné také z: <http://ieeexplore.ieee.org/document/6696595/>
- [10] OLSON, Edwin. AprilTag: A robust and flexible visual fiducial system. *2011 IEEE International Conference on Robotics and Automation. IEEE, 2011, 2011, , 3400-3407*. DOI: 10.1109/ICRA.2011.5979561. ISBN 978-1-61284-386-5. Dostupné také z: <http://ieeexplore.ieee.org/document/5979561/>
- [11] Camera suite [online]. APRIL ROBOTICS LABORATORY. 30. 5. 2016 [cit. 2018-03-26]. Dostupné z: https://april.eecs.umich.edu/wiki/Camera_suite
- [12] NISTER, D. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004, 26(6), 756-770. DOI: 10.1109/TPAMI.2004.17. ISSN 0162-8828. Dostupné také z: <http://ieeexplore.ieee.org/document/1288525/>

- [13] ARUN, K. S., T. S. HUANG a S. D. BLOSTEIN. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1987, PAMI-9(5), 698-700. DOI: 10.1109/TPAMI.1987.4767965. ISSN 0162-8828. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4767965>
- [14] KNEIP, Laurent, Davide SCARAMUZZA a Roland SIEGWART. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. CVPR 2011. IEEE, 2011, 2011, , 2969-2976. DOI: 10.1109/CVPR.2011.5995464. ISBN 978-1-4577-0394-2. Dostupné také z: <http://ieeexplore.ieee.org/document/5995464/>
- [15] FRAUNDORFER, Friedrich a Davide SCARAMUZZA. Visual Odometry: Part II. 2012, 19(2), 78-90. DOI: 10.1109/MRA.2012.2182810. ISSN 1070-9932. Dostupné také z: <http://ieeexplore.ieee.org/document/6153423/>
- [16] HORÁK, Karel. Popis objektů. FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ VUT V BRNĚ. Počítačové vidění [online]. Brno, 2008, s. 89-100 [cit. 2018-04-06]. Dostupné z: http://www.uamtold.feec.vutbr.cz/vision/TEACHING/MPOV/Pocitacove_videni_S.pdf
- [17] HAVLE, Otto. Přehled trhu: inteligentní kamery. *Automa*. 2009(05), 22-23. Dostupné také z: http://automa.cz/Aton/FileRepository/pdf_articles/39009.pdf
- [18] Feature Detection and Description. OpenCV 3.0.0-dev documentation [online]. 10. 11. 2014 [cit. 2018-04-06]. Dostupné z: https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_table_of_contents_feature2d/py_table_of_contents_feature2d.html
- [19] HARRIS, Chris a Mike STEPHENS. A COMBINED CORNER AND EDGE DETECTOR. 1988. Dostupné také z: <http://www.bmva.org/bmvc/1988/avc-88-023.pdf>
- [20] LOWE, David G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. 2004, 60(2), 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94. ISSN 0920-5691. Dostupné také z: <http://link.springer.com/10.1023/B:VISI.0000029664.99615.94>
- [21] BARBORKA, Petr. Analýza metod pro detekci příznaků v digitalizovaném obraze [online]. 2016 [cit. 2018-04-09]. Dostupné z: <http://hdl.handle.net/11025/23781>. Bakalářská práce. Západočeská univerzita v Plzni. Vedoucí práce Petr Neduchal.
- [22] BAY, Herbert, Tinne TUYTELAARS a Luc VAN GOOL. SURF: Speeded Up Robust Features. *Computer Vision – ECCV 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, 2006, , 404-417. Lecture Notes in Computer Science. DOI: 10.1007/11744023_32. ISBN 978-3-540-33832-1. Dostupné také z: http://link.springer.com/10.1007/11744023_32
- [23] ROSTEN, Edward a Tom DRUMMOND. Machine Learning for High-Speed Corner Detection. *Computer Vision – ECCV 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, 2006, , 430-443. Lecture Notes in Computer Science. DOI: 10.1007/11744023_34. ISBN 978-3-540-33832-1. Dostupné také z: http://link.springer.com/10.1007/11744023_34
- [24] CALONDER, Michael, Vincent LEPETIT, Christoph STRECHA a Pascal FUA. BRIEF: Binary Robust Independent Elementary Features. *Computer Vision – ECCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, 2010, , 778-792. Lecture Notes in Computer Science. DOI:

- 10.1007/978-3-642-15561-1_56. ISBN 978-3-642-15560-4. Dostupné také z: http://link.springer.com/10.1007/978-3-642-15561-1_56
- [25] RUBLEE, Ethan, Vincent RABAUD, Kurt KONOLIGE a Gary BRADSKI. ORB: An efficient alternative to SIFT or SURF. 2011 International Conference on Computer Vision. IEEE, 2011, 2011, , 2564-2571. DOI: 10.1109/ICCV.2011.6126544. ISBN 978-1-4577-1102-2. Dostupné také z: <http://ieeexplore.ieee.org/document/6126544/>
- [26] MUJA, Marius a David G. LOWE. Scalable Nearest Neighbor Algorithms for High Dimensional Data. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014, 36(11), 2227-2240. DOI: 10.1109/TPAMI.2014.2321376. ISSN 0162-8828. Dostupné také z: <http://ieeexplore.ieee.org/document/6809191/>
- [27] GEIGER, A., P. LENZ a R. URTASUN. Are we ready for autonomous driving? The KITTI vision benchmark suite. 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, 2012, , 3354-3361. DOI: 10.1109/CVPR.2012.6248074. ISBN 978-1-4673-1228-8. Dostupné také z: <http://ieeexplore.ieee.org/document/6248074/>
- [28] CVIŠIĆ, Igor, Josip ĆESIĆ, Ivan MARKOVIĆ a Ivan PETROVIĆ. SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. Journal of Field Robotics. DOI: 10.1002/rob.21762. ISSN 15564959. Dostupné také z: <http://doi.wiley.com/10.1002/rob.21762>
- [29] CVISIC, Igor a Ivan PETROVIC. Stereo odometry based on careful feature selection and tracking. 2015 European Conference on Mobile Robots (ECMR). IEEE, 2015, 2015, 1-6. DOI: 10.1109/ECMR.2015.7324219. ISBN 978-1-4673-9163-4. Dostupné také z: <http://ieeexplore.ieee.org/document/7324219/>
- [30] GEIGER, Andreas, Julius ZIEGLER a Christoph STILLER. StereoScan: Dense 3d reconstruction in real-time. 2011 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011, 2011, , 963-968. DOI: 10.1109/IVS.2011.5940405. ISBN 978-1-4577-0890-9. Dostupné také z: <http://ieeexplore.ieee.org/document/5940405/>
- [31] BUCZKO, Martin a Volker WILLERT. Flow-decoupled normalized reprojection error for visual odometry. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2016, 2016, , 1161-1167. DOI: 10.1109/ITSC.2016.7795703. ISBN 978-1-5090-1889-5. Dostupné také z: <http://ieeexplore.ieee.org/document/7795703/>
- [32] SCARAMUZZA, Davide, Friedrich FRAUNDORFER a Roland SIEGWART. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. 2009 IEEE International Conference on Robotics and Automation. IEEE, 2009, 2009, , 4293-4299. DOI: 10.1109/ROBOT.2009.5152255. ISBN 978-1-4244-2788-8. Dostupné také z: <http://ieeexplore.ieee.org/document/5152255/>
- [33] CHOI, Sunglok, Jaehyun PARK a Wonpil YU. Simplified epipolar geometry for real-time monocular visual odometry on roads. International Journal of Control, Automation and Systems. 2015, 13(6), 1454-1464. DOI: 10.1007/s12555-014-0157-6. ISSN 1598-6446. Dostupné také z: <http://link.springer.com/10.1007/s12555-014-0157-6>
- [34] BOULEKCHOUR, Mohammed a Nabil AOUF. Robust motion estimation using covariance intersection. 22nd Mediterranean Conference on Control and Automation. IEEE, 2014, 2014, ,

1014-1019. DOI: 10.1109/MED.2014.6961507. ISBN 978-1-4799-5901-3. Dostupné také z: <http://ieeexplore.ieee.org/document/6961507/>

[35] FORSTER, Christian, Zichao ZHANG, Michael GASSNER, Manuel WERLBERGER a Davide SCARAMUZZA. SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems. IEEE Transactions on Robotics. 2017, 33(2), 249-265. DOI: 10.1109/TRO.2016.2623335. ISSN 1552-3098. Dostupné také z: <http://ieeexplore.ieee.org/document/7782863/>

[36] FORSTER, Christian, Matia PIZZOLI a Davide SCARAMUZZA. SVO: Fast semi-direct monocular visual odometry. 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2014, 2014, , 15-22. DOI: 10.1109/ICRA.2014.6906584. ISBN 978-1-4799-3685-4. Dostupné také z: <http://ieeexplore.ieee.org/document/6906584/>

[37] SMUTNÝ, Vladimír. *Robotika: Popis polohy tělesa*. Dostupné také z: <http://cmp.felk.cvut.cz/cmp/courses/ROB/roblec/geometry-notecz.pdf>

[38] Intel® Core™ i7-2630QM Processor [online]. © INTEL CORPORATION. [cit. 2018-04-30]. Dostupné z: https://ark.intel.com/products/52219/Intel-Core-i7-2630QM-Processor-6M-Cache-up-to-2_90-GHz

[39] AcA1920-155uc - Basler ace [online]. BASLER AG. 2018 [cit. 2018-04-30]. Dostupné z: <https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1920-155uc/>

Příloha A – obsah CD:

Přiložené CD obsahuje elektronickou verzi této práce (Diplomová práce.pdf) a zdrojové kódy programu s implementovanou metodou pro vizuální odometrii a automatickým testem příznaků ve složce „Visual odometry“.