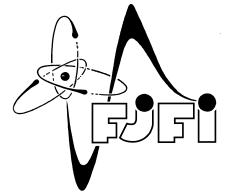




ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta jaderná a fyzikálně inženýrská



Automatická registrace snímků video

Automatic registration of video frames

Bakalářská práce

Autor: **Daniela Lichnovská**

Vedoucí práce: **doc. Ing. Filip Šroubek, Ph.D. DSc.**

Konzultant: **Ing. Lukáš Tryner**

Akademický rok: **2020/2021**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: Daniela Lichnovská
Studijní program: Aplikace přírodních věd
Studijní obor: Matematické inženýrství
Studijní zaměření: Aplikované matematicko-stochastické metody
Název práce (česky): Automatická registrace snímků videa
Název práce (anglicky): Automatic registration of video frames

Pokyny pro vypracování:

- 1) Seznamte se základními postupy registrace obrazů a skládání panoramat.
- 2) Navrhněte registrační metodu vhodnou na skládání snímků videa zachycující regál v obchodě.
- 3) Navrhněte metodu na detekci kvality registrace.
- 4) Analyzujte vliv rozmazání způsobené pohybem ruky na kvalitu registrace.
- 5) Navrhněte metodu na výběr snímků z videa optimalizující prostorové pokrytí regálu a kvalitu registrace.
- 6) Experimentálně ověřte navržený postup na reálných videích.
- 7) V případě, že bude k dispozici metoda na detekci zboží, tak ověřte funkčnost navrženého postupu i s ohledem na schopnost detektovat správně zboží.

Doporučená literatura:

- 1) R. C. Gonzales, R. E. Woods, Digital Image Processing (2nd ed.). Prentice Hall, 2002.
- 2) B. Zitová, J. Flusser, Image registration methods: a survey. Image and Vision Computing 21, 2003, 977-1000.
- 3) J. Mustaniemi, J. Kannala, S. Särkkä, J. Matas, J. Heikkilä, Fast motion deblurring for feature detection and matching using inertial measurements. In 'IEEE ICPR', Beijing, China, 2018.

Jméno a pracoviště vedoucího bakalářské práce:

doc. Ing. Filip Šroubek, Ph.D. DSc.
ÚTIA AV ČR Praha, v.v.i., Pod Vodárenskou věží 4, 182 00 Praha 8

Jméno a pracoviště konzultanta:

Ing. Lukáš Tryner
DataSentics, a.s., Washingtonova 1599/17 , Praha 1

Datum zadání bakalářské práce: 31.10.2020

Datum odevzdání bakalářské práce: 7.7.2021

Doba platnosti zadání je dva roky od data zadání.

V Praze dne 30.10.2020

RB

garant oboru

R. Hlaváč

vedoucí katedry



děkan

Poděkování:

Chtěl bych zde poděkovat především svému školiteli za pečlivost, ochotu, vstřícnost a od-
borné i lidské zázemí při vedení mé diplomové práce. Dále děkuji svému konzultantovi za
.....

Čestné prohlášení:

Prohlašuji, že jsem tuto práci vypracoval samostatně a uvedl jsem všechnu použitou literaturu.

V Praze dne 30. března 2021

Daniela Lichnovská

Název práce:

Název práce

Autor: Daniela Lichnovská

Obor: Matematické inženýrství

Zaměření: Aplikované matematicko-stochastické metody

Druh práce: Bakalářská práce

Vedoucí práce: doc. Ing. Filip Šroubek, Ph.D. DSc., ÚTIA AV ČR Praha, v.v.i., Pod Vodárenskou věží 4, 182 00 Praha 8

Konzultant: Ing. Lukáš Tryner, DataSentics, a.s., Washingtonova 1599/17, Praha 1

Abstrakt: Abstrakt max. na 10 řádků. Abstrakt max. na 10 řádků.

Klíčová slova: klíčová slova (nebo výrazy) seřazená podle abecedy a oddělená čárkou

Title:

Automatic registration of video frames

Author: Daniela Lichnovská

Abstract: Max. 10 lines of English abstract text. Max. 10 lines of English abstract text.

Key words: keywords in alphabetical order separated by commas

Obsah

Úvod	7
1 Registrace obrazu	8
1.0.1 co to je	8
1.1 Detekce příznaků	8
1.1.1 Metoda SURF	8
1.1.2 Extrakce příznaků	8
1.2 Korespondence příznaků	8
1.3 Estimace modelu transformace	8
1.3.1 Affiní tr	8
1.3.2 Projektivní tr	8
1.4 resampling	8
1.5 pipeline	8
2 Výběr snímků z videa	9
2.1 focus measure	9
2.2 Velikost okna	9
2.3 Error threshold	9
2.4 pipeline	10
3 Registrace panoramatu	11
3.1 computeTforms	11
3.2 computeLimits	11
3.3 createPanorama	11
3.4 Omezení	11
4 Experimenty	12
4.1 Opakovatelnost(repeatability) a loc error	12
4.2 Vliv rozmazání na registraci	13
4.3 vliv artefaktů	13
5 Detekce cenovek	14
5.1 Omezení	14
5.2 Výsledky	14
Závěr	16

Úvod

Text úvodu....

Kapitola 1

Registrace obrazu

1.0.1 co to je

Registrace obrazu je proces kombinování dvou a více snímků pořízených z odlišných pohledů, času nebo zdrojů do jednoho současného systému [3]. Je například často používána při zpracovávání dat ze satelitů, v počítačovém vidění nebo medicíně.

Registraci lze rozdělit do čtyř kroků:

1. detekce příznaků
2. korespondence příznaků
3. odhad transformace
4. transformace snímku

1.1 Detekce příznaků

1.1.1 Metoda SURF

1.1.2 Extraktce příznaků

1.2 Korespondence příznaků

1.3 Estimace modelu transformace

1.3.1 Affiní tr

1.3.2 Projektivní tr

1.4 resampling

1.5 pipeline

Kapitola 2

Výběr snímků z videa

2.1 focus measure

[<https://www.pyimagesearch.com/2015/09/07/blur-detection-with-opencv/>]

Nejkvalitnější snímky z videa jsou ty s nejmenším rozmazáním. Méně rozmazaný snímek zároveň zřejmě bude mít výraznější hrany než ten více rozmazaný. Hrany detekujeme pomocí konvoluce Laplaceova operátoru s vybraným snímkem. Výsledný (absolutní) průměr této matice nám bude sloužit jako nástroj k porovnávání rozmazání mezi snímky. Tato metoda není úplně přesná, protože se scéna na snímcích konstantě mění, nicméně se nemění tak dramaticky, aby nebyla použitelná.

2.2 Velikost okna

Předpokládáme rychlosť kamery 20 cm/s a frame rate 30FPS , tedy posun 2/3 cm/frame. Chceme aby se dva snímky překrývali alespon ze 1/3 (při menším překryvu by počítání tr. mat mohlo být obtížné/nepřesné) a zároveň se nepřerývali více než ze 1/2 (minimalizace počtu snímků). Předpokládáme, že snímek zobrazuje vertikálně 1 m ve skutečnosti. Pak se dostaváme na hodnotu okna 37-75 framů.

Nápad : Počítání rychlosti kamery. 2 kola

1) Vyberu dostatečně ostré snímky (jen na začátku/ celé video?). Možná budu potřebovat pevné okno? 2) Spočítám si projektivní tr. matice. 3) Pomocí velikosti posunu tr. matice a časovou vzdáleností spočítám průměrnou rychlosť. 4) Přes známou rychlosť spočítám přesnější okno

100 frameu uprostred-spočítam rychlosť z toho dynamicke zvetsivano/zmensovani okna podle predchoziholepsi

2.3 Error threshold

Pro odhad dobré transformace použijeme následující metodu: Označíme smer. odchylku intenzity snímků jako errThresh. Tato hodnota nám bude sloužit jako prah při porovnávání s chybou transformace. Ta je spočítána jako $\text{sum}(\text{abs}(I1-I2))$ kde I1,I2 jsou překrývající se části snímků. Pokud se chyba žádného snímků nevejde pod ErrThresh, je vybrán snímek s nejmenší chybou transformace v daném okně.

$$\text{mean} = \text{sum}(x)/\text{length}(x)$$

$$\text{variance} = \text{sum}((x - \text{mean}(x))^2)/(\text{length}(x) - 1)$$

$$\text{std} = \sqrt{\text{variance}}$$

2.4 pipeline

,Pro všechny snímky ve videu je spočítána focus measure. Všechny ostatní práce se snímky je postupně dělána ve vyhrazených oknech, což je soubor snímků vzdálený od předchozího vybraného snímku minimální a maximální vzdáleností(vzdáleností je myšleno časové pořadí snímků). Pro první průchod je jednoduše vybrán snímek s největší fm a spočítán ErrThresh pomocí něj. Zbytek cyklu probíhá takto:

1) Všechny snímky okna jsou v sestupném pořadí seřazeny podle jejich focus measure.

2) Snímky jsou jeden po druhém registrovány na vybraný snímek z předchozího okna, dokud není nalezen snímek s chybou transformace menší než ErrThresh. Pokud ani jeden není menší, je vybrán snímek s nejmenší chybou tr.

3) Tento snímek je přidán do konečného výběru snímků

Kapitola 3

Registrace panoramu

[<https://www.mathworks.com/help/vision/ug/feature-based-panoramic-image-stitching.html>]

3.1 computeTforms

Transformační matici jsou počítány následující způsobem: Snímek je převeden do černobílé (1 vrstva místo 3 u RGB) a jsou na něm detekovány body pomocí SURF detektora. Z těchto bodů jsou vybrány příznaky, které jsou následně spárovány s příznaky z předchozího snímku. Tyto 2 seznamy jsou poté použity funkcí estimateGeometricTransform, která spočítá odpovídající transformaci. Pro potřeby panoramu je použita affinní transformace, protože u projektivní dochází při součinu matic ke velké chybě.

Pokud chceme, aby referenčním snímkem v panoramu byl jiný než první, je potřeba je přepočítat pomocí $tforms(j).T = tforms(j).T * Tinv.T$; kde $Tinv$ je inverze ref. snímku.

3.2 computeLimits

Je třeba znát rozměry výsledného panoramu, čehož docílíme transformací 4 bodů okrajů u všech snímku a následným výběrem těch nejzaších. V tomto kroce je dobré ověřit velikost pro výsledné panorama, pokud je příliš velká, evidentně někde nastala chyba.

3.3 createPanorama

Díky znalosti rozměrů panoramu si můžeme vytvořit konečný souřadný systém, do kterého nejdříve vložíme masku (opět 4 kraje původního snímku) a následně do místa masky vložíme transformovaný snímek. (Masku pak můžeme využít na zobrazení vizualizaci poskládání všech snímků)

3.4 Omezení

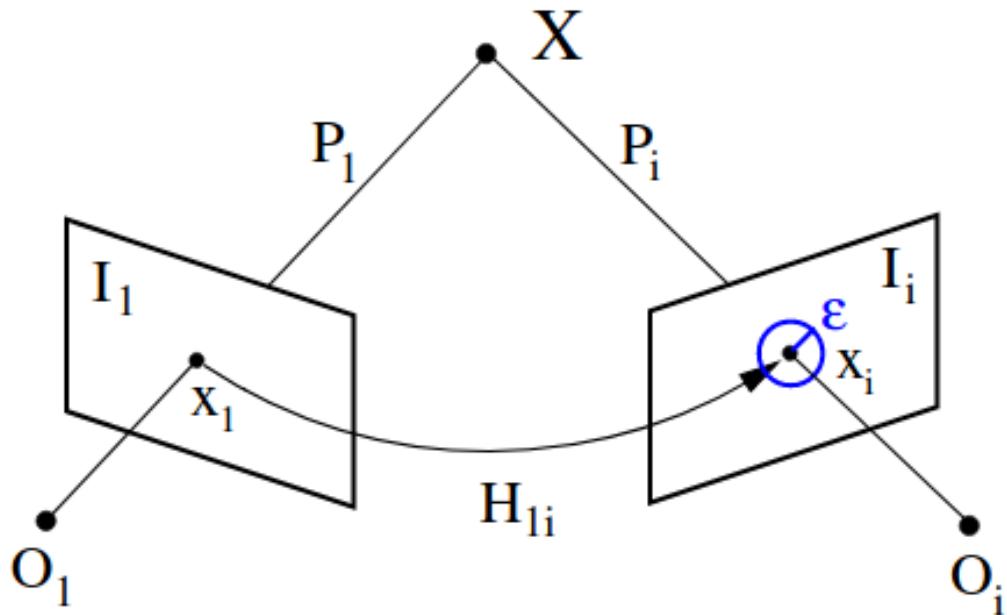
Jak bylo výše uvedeno, u dlouhých panoram s jakými pracujeme my, je třeba použít affiní, či podobnostní transformaci, neboť při součinu projektivních transformací dochází ke stále větší a větší chybě, a při více než 5-7 snímcích ani nelze žádné panorama vytvořit.

Kapitola 4

Experimenty

4.1 Opakovatelnost(repeatability) a loc error

Opakovatelnost [repeat] Repeatability rate definujeme jako počet stejných bodů vyskytujících se na



Obrázek 4.1: Ilustrace principu opakovatelnosti

obou snímcích vzhledem k celkovém počtu detekovaných bodů. Z detekovaných bodů vybíráme pouze ty, které jsou součástí scény vyskytujících se na obou snímcích.

Je důležité vzít v úvahu nepřesnost detekce. Korespondující bod se nebude nacházet přesně daném místě určeném maticí, ale v jeho eps -okolí. eps -repeatabilitu definujeme jako:

$$Ri(\text{eps}) = \{(x_i, x_j) | dist(H_{ij}x_i, x_j) < \text{eps}\}$$

Počet detekovaných bodů ve společné části snímku nebude stejný, proto vybereme menší z nich, abychom zajistili existenci bodu z eps -repeatability. Repeatability rate je tedy definováno následovně:

$$ri(\text{eps}) = \text{abs}(Ri(\text{eps}))/\min(n_i, j)$$

4.2 Vliv rozmazání na registraci

Pro sledování vlivu rozmazání na kvalitu registrace použijeme dvojice sousedících snímků, které uměle rozmažeme aplikací motion-blur filtru náhodné orientace v rozsahu $<0,190>$ a pevně zvolené délky v rozsahu $<1,50>$, kdy je postupně přidávána po inkrementech 5(pixelů?). Z jedné dvojice tak získáme 11 (12 s originálními) páru rozmazaných snímků. U nich pak najdeme příznaky, namatchujeme je a odhadneme homografní matici. Pro E v Eps-repeatabilite dosadíme 10. Takte získáme repeatability rate.

Dále chceme znát localization error, tedy průměrnou vzdálenost bodu predikovaného tr. maticí od místa, kde se bod skutečně nachází. To se dá vyjádřit jako:

$$dist(i) = \text{norm}(p2(i,:) - \text{predicted}(i,:)); \text{mean}(dist)$$

Takto projdeme celý daný set snímků. Tady je graf setu 1:

Na vertikální ose je repeatability rate a jeho rozptyl, na horizontální délka filtru. Jak lze vidět z grafu, do cca 30 pixelu je kvalita registrace poměrně spolehlivá, pak ale začne rychle klesat.

4.3 vliv artefaktů

Sledování vlivu zmenšování datového toku kvalitu registrace. Plánovaný postup:

- 1)Vygeneruju set videí se různým datovým tokem
- 2)Problém s výběrem snímku (od nějakého to už nezvládne vybrat snímek)
řešení: vybírám předem dané snímky (znám čísla z originálu)
- 3>Dále postup stejný jako u bluru (detekce příznaků, matching, estimateGeometricTransform,etc...))

Kapitola 5

Detekce cenovek

V této části si ukážeme využití výše zmíněných metod na problému počítání cenovek. V podstatě jde o schopnost detekce stejného objektu na více snímcích a následném rozdělení tak, aby každý snímek měl "své" cenovky. Pro toto je využití vlastností transformačních matic ideální.

Pro samotnou detekci cenovek byla využita typ sítě YOLOv5[<https://github.com/ultralytics/yolov5>] natrénována na cenovky od firmy DataSentics. Ta vyniká především pro svou rychlosť i malou velikost, čímž je vhodná pro využití v mobilních zařízeních. Vstupem je snímek a výstupem list souřadnic bounding boxů.

Snímky jsou získány z videa postupem výše uvedeném. Mezi těmito snímkami je spočítány matice projektivní transformace a je také (odděleně) vygenerováno panorama pomocí affiní transformace. Toto panorama slouží k ručnímu sčítání cenovek.

V každém cyklu jsou vzaty 3 po sobě jsoucí snímky, v nichž jsou následně detekovány cenovky. První a třetí bboxy jsou převedeny transformací do souřadného systému druhého snímku. Poté jsou přes Python knihovnu Shapely[] souřadnice bboxů převedeny do polygonů pro snadnější manipulaci.

Polygony prvního a druhého snímku jsou dávány přes sebe a pokud je nalezen překryv, polygon je odstraněn z listu ve 3. a zůstává pouze ve 1. Následně jsou podobným způsobem kontrolovány polygony prvního a druhého a třetího a druhého. Při nalezení překryvu ale o tom, ze kterého listu bude polygon smazán rozhoduje velikost obsahu cenovky.

Po prvním cyklu neprochází již detekované snímky sítí znova, místo toho jsou brány seznamy z předchozích 2 cyklů. Díky tomu nedochází k duplikaci cenovek na různých snímcích. Mohlo by v nějakých případech dojít k duplikaci, pokud by se objekt objevil na 4 a více snímcích po sobě, nicméně k tomu v podstatě nedochází, vzhledem k tomu že snímky jsou vybrány z videa, tak, aby neměly tak velký překryv. Nehledě na to, že při součinu projektivních transformačních matic se rychle kumuluje chyba po několika cyklech by se bboxy přestaly překrývat.

5.1 Omezení

Sít má problém s cenovkami velmi blízko u sebe a najde neexistující cenovky. Také má problém správně detekovat cenovky vyloženě přes sebe a neobvyklých formátů (elektronické, moc dlouhé)

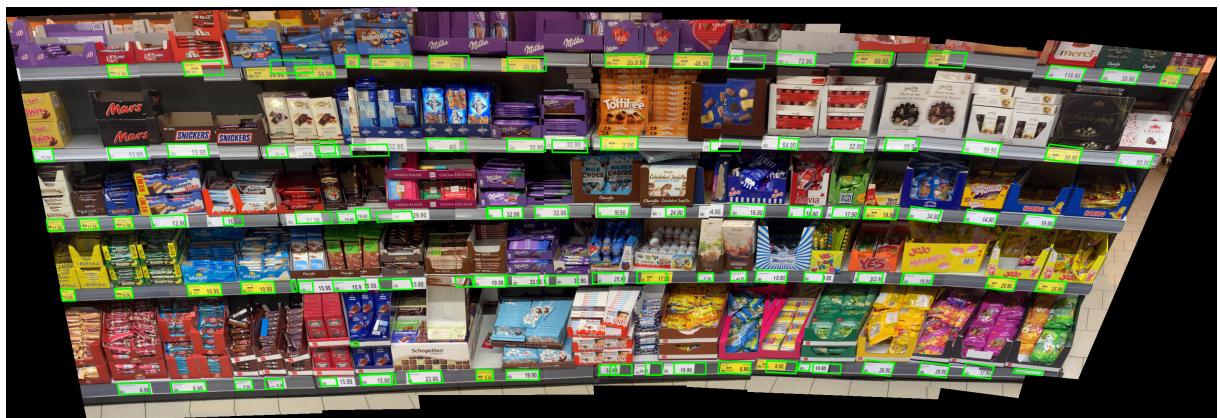
5.2 Výsledky

Číslo setu	1	2	3	4	5
Sečteno ručně	188	190	xxx	xxx	xxx
Sečteno programem	102	96	xxx	xxx	xxx

Tabulka 5.1: TABLETABLETABLE.



Obrázek 5.1: Zobrazení cenovek na panoramatu z vid30.



Obrázek 5.2: A boat.

Závěr

Text závěru....

Bibliografie

- [1] abidrahmak(github). *Introduction to SIFT (Scale-Invariant Feature Transform)*. 2013. URL: https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html.
- [2] Janne Mustaniemi et al. “Fast Motion Deblurring for Feature Detection and Matching Using Inertial Measurements”. In: (květ. 2018).
- [3] Barbara Zitová a Jan Flusser. “Image Registration Methods: A Survey”. In: *Image and Vision Computing* 21 (říj. 2003), s. 977–1000. doi: [10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9).

Seznam obrázků

4.1	Zobrazení	12
5.1	Zobrazení cenovek na panoramatu z vid30.	15
5.2	A boat.	15

Seznam tabulek

5.1 TABLETABLETABLE	15
-------------------------------	----