Vizualni SLAM z gostim ujemalnikom značilk RoMa

Avtor: Matic Šinigoj, Mentor: prof. dr. Danijel Skočaj, Somentor: asist. dr. Matej Dobrevski
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

1. Uvod in Motivacija

Sočasna lokalizacija in kartiranje (SLAM) je ključna tehnologija za avtonomne sisteme, kot so roboti in avtonomna vozila. Vizualni SLAM (vSLAM), ki za navigacijo uporablja izključno kamere, postaja vse bolj pomemben zaradi velike količine informacij, ki jih kamere zajamejo in njihove dostopnosti.

Motivacija za to nalogo izhaja iz hitrega napredka na področju globokega učenja, ki ponuja zmogljivejše alternative klasičnim, ročno zasnovanim metodam v računalniškem vidu. Tradicionalni sistemi SLAM, kot je ORB-SLAM2 [7], se zanašajo na redke značilke, kar jih omejuje v okoljih z malo teksture ali pri hitrem gibanju. Po drugi strani pa moderni, na učenju temelječi pristopi, kot so gosti ujemalniki značilk, obljubljajo bistveno večjo robustnost in natančnost.

Osrednji cilj te naloge je bil zasnovati, implementirati in ovrednotiti nov, hibridni sistem za vizualni SLAM, ki združuje preverjeno geometrijsko ogrodje klasičnih pristopov z najsodobnejšimi komponentami globokega učenja. Poudarek je na integraciji:

- **Gostega ujemalnika značilk RoMa** [1] za robustno iskanje korespondenc med slikami.
- **Sistema za prepoznavanje lokacij AnyLoc** [2] za učinkovito zaznavanje in zapiranje zank.

2. Arhitektura in Implementacija Sistema

Implementiran sistem temelji na modularni arhitekturi, ki je jasno razdeljena na sledilni in optimizacijski del.

Sledilni del (Front-end)

Njegova glavna naloga je obdelava vhodnih slik (ključnih sličic) in sprotno ocenjevanje gibanja kamere ter gradnja 3D zemljevida.

1. **Izbor ključnih sličic:** Sistem ne obdeluje vsake slike, temveč le tiste, ki prinašajo dovolj novih informacij, določenih na podlagi časovnega intervala in mediane paralakse.
2. **Ujemanje značilk:** Za iskanje korespondenc med ključnimi sličicami skrbi **RoMa**, ki vrne gosto polje ujemanj. Iz tega polja z vzorčenjem izberemo najbolj zanesljive pare.
3. **Inicializacija in sledenje:** Sistem se inicializira z oceno **esencialne matrike** [5] iz prvih dveh pogledov. Nadaljnje sledenje poteka z algoritmom **PnP (Perspective-n-Point)** [3], ki lego kamere ocenjuje na podlagi 2D-3D korespondenc med trenutno sliko in obstoječim 3D zemljevidom.
4. **Lokalna optimizacija:** Za sprotno odpravljanje manjših napak sistem periodično izvaja **snopovno poravnavo (Bundle Adjustment)** [9] na drsečem oknu zadnjih nekaj ključnih sličic in opazovanih 3D točk.

Optimizacijski del (Back-end)

Ta del je odgovoren za dolgoročno konsistentnost zemljevida in trajektorije.

1. **Zapiranje zank:** Ko se kamera vrne na že obiskano lokacijo, sistem **AnyLoc** prepozna podobnost med slikama. Ta dogodek (zanka) se nato geometrijsko preveri in v faktorski graf doda kot močna omejitev med dvema časovno oddaljenima deloma poti.
2. **Globalna optimizacija:** Po zaznavi zanke ali na koncu izvajanja se sproži globalna optimizacija celotnega **faktorskega grafa**. Z Levenberg-Marquardtovim algoritmom [4, 6] se popravijo vse lege kamer in 3D točke, s čimer se odpravi nakopičeni zdrs (drift).

3. Rezultati in Analiza

Sistem smo ovrednotili na standardni podatkovni množici **TUM RGB-D** [8] z uporabo mere ATE (absolutna napaka trajektorije) [8]. Sistem se je izkazal za zelo robustnega, saj je uspešno obdelal vse testne sekvence.

Analiza je pokazala, da ima **lokalna optimizacija** ključen vpliv na natančnost. Na slikah 1 in 2 je prikazana vizualna primerjava delovanja sistema na sekvenci `freiburg1_desk`. Ko je lokalna optimizacija vključena, se ocenjena trajektorija bistveno bolj prilega referenčni, kar potrjuje, da ta mehanizem učinkovito zmanjšuje sprotno kopičenje napak.

Table 1: Končna absolutna napaka trajektorije (ATE RMSE [m]) na vseh sekvencah `freiburg1` po globalni optimizaciji.

Sekvenca	ATE (m)
`freiburg1_360`	0.164
`freiburg1_desk`	0.035
`freiburg1_desk2`	0.753
`freiburg1_floor`	0.413
`freiburg1_room`	0.603
`freiburg1_rpy`	0.049
`freiburg1_xyz`	0.027

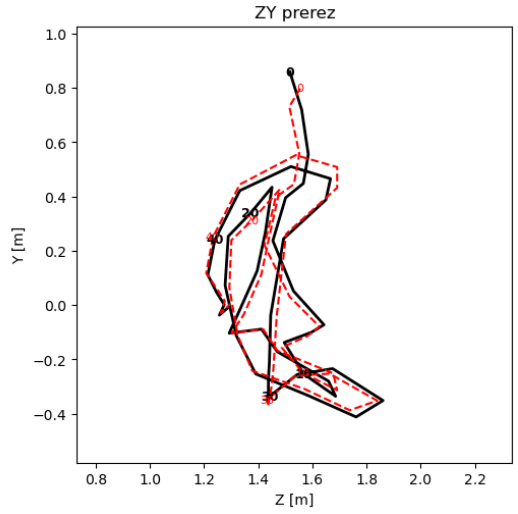

Figure 1: Primerjava končne trajektorije (rdeče) z referenčno (črno) na sekvenci `desk` **z vklopljeno** lokalno optimizacijo.

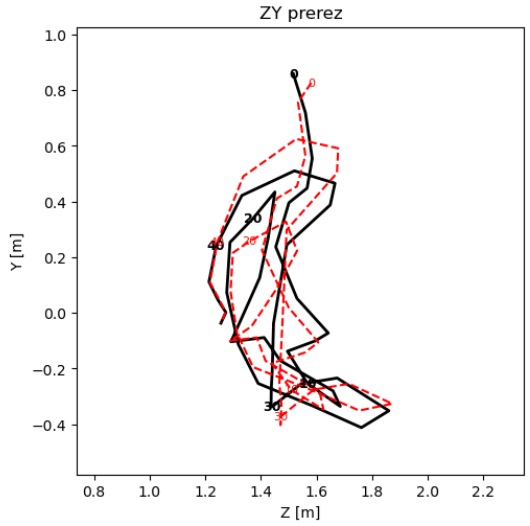

Figure 2: Primerjava trajektorije (rdeče) z referenčno (črno) na sekvenci `desk` **z izklopljeno** lokalno optimizacijo. Opazen je večji zdrs.

V tabeli 1 so zbrani končni rezultati delovanja celotnega sistema (z vključenimi vsemi optimizacijami) na vseh testiranih sekvencah. Rezultati potrjujejo, da sistem dosega konkurenčno natančnost v različnih scenarijih, od preprostih premikov do kompleksnejših poti z zapiranjem zank.

4. Sklep

V nalogi smo uspešno zasnovali in implementirali nov, hibridni sistem za vizualni SLAM, ki dokazuje potencial integracije modernih, na učenju temelječih komponent v klasična ogrodja. Sistem je robusten, dosega konkurenčno natančnost in predstavlja trdno osnovo za nadaljnje raziskave. Glavni izzivi za prihodnje delo ležijo v razvoju bolj prilagodljive strategije za izbor ključnih sličic in optimizaciji hitrosti za delovanje v realnem času.

Literatura

[1] Johan Edstedt et al. “RoMa: Robust Dense Feature Matching”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024.

[2] Nikhil Keetha et al. “AnyLoc: A Foundation Model for Visual Place Recognition”. In: *arXiv preprint arXiv:2311.17258* (2023).

[3] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. “EPnP: An accurate O(n) solution to the PnP problem”. In: *International journal of computer vision* 81 (2009), pp. 155–166.

[4] Kenneth Levenberg. “A method for the solution of certain non-linear problems in least squares”. In: *Quarterly of applied mathematics* 2.2 (1944), pp. 164–168.

[5] H Christopher Longuet-Higgins. “A computer algorithm for reconstructing a scene from two projections”. In: *Nature* 293.5828 (1981), pp. 133–135.

[6] Donald W Marquardt. “An algorithm for least-squares estimation of nonlinear parameters”. In: *Journal of the society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.

[7] Raul Mur-Artal and Juan D. Tardos. “ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras”. In: *IEEE Transactions on Robotics* 33.5 (2017), pp. 1255–1262.

[8] Jürgen Sturm et al. “A benchmark for the evaluation of RGB-D SLAM systems”. In: *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. 2012, pp. 573–580.

[9] Bill Triggs et al. “Bundle adjustment—a modern synthesis”. In: *International workshop on vision algorithms*. Springer. 1999, pp. 298–372.