Dinamička analiza 3D scena

analiza kretanja u slikovnoj ravnini

Siniša Šegvić i Zoran Kalafatić

UniZg-FER ZEMRIS

Uvod

Dinamička analiza scena: podskup računalnog vida u kojem je razumijevanje scene *kontinuirani* proces

Česti zadatci:

- nepokretna kamera, pokretni objekti, nepokretna pozadina
- pokretna kamera, nepokretna scena
- pokretna kamera, pokretni objekti, nepokretna pozadina

Izazovi s kojima se susrećemo:

- računska složenost, memorijski zahtjevi i stvarno vrijeme
- osvjetljenje, šum, ekspozicija, otvor objektiva
- nesavršenost opreme (rolling shutter, interlacing, izobličenja)
- scenario (nedostatak teksture, aliasiranje, pokretni objekti)

Uvod (2)

Kontekst (1):

- računalni vid
  - zadatci:
    - pronalaženje (objekti ili istaknute točke),
    - asociranje (uparivanje ili praćenje),
    - geometrija više pogleda
  - metode: rekonstrukcija i klasifikacija
  - danas sve metode rješenje pronalaze optimizacijom!
- obrada slike: filtriranje, transformiranje, višerezolucijska analiza
- strojno učenje: klasifikacija, regresija

Uvod (3)

Kontekst (2):

- MATEMATIKA: algebra, analiza, probabilistički filtri, stohastički procesi, ...
- STATISTIKA: robusna procjena parametara, modeliranje razdioba, klasifikacija, ...
- programsko inženjerstvo
- umjetna inteligencija

Sadržaj:

- Uvod
- Motivacija: primjene dinamičke analize 3D scena
- Određivanje korespondencija u slikama
  - pronalaženje značajki (Harris)
  - analiza 2D kretanja (KLT)
  - stereoskopska rekonstrukcija (SGM, MC-CNN)
- Robusna estimacija

Primjene dinamičke analize scena

Najčešće pretpostavke:

- pokretna kamera (ili više kamera)
- *najčešće* vidljiv *znatan* dio nepokretne scene
- pokretni objekti otežavaju analizu
  (najčešće nisu interesantni *per se*)
- problemi: promjenljivo osvjetljenje, scene bez značajki, skalabilnost, pokretni objekti, pogreške pri detekciji i asocijaciji

Elementi rješenja:

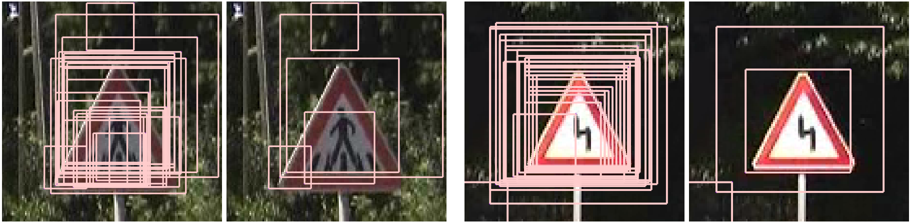
- 2D analiza (pronalaženje, uparivanje, praćenje, raspoznavanje)
- interni prikaz scene (statičan ili dinamičan)
- 3D interpretacija (kretanje kamere, struktura scene)

Dinamička analiza 3D scena: znakovi 2010

Primjer: poboljšanje lokacijske točnosti detekcije praćenjem

Detekcija trokutnih znakova u pojedinačnim slikama:

- **izvrstan odziv**: preko 90% znakova je detektirano
- nedovoljna preciznost (89%)
- neprecizna lokalizacija

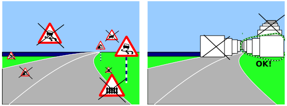


Dinamička analiza 3D scena: znakovi 2010

Ideja: poboljšati rezultate integracijom odziva kroz veći broj slika

- pronaći konzistentnu trajektoriju s najvećim brojem potvrda

- zahtijevati geometrijsku prikladnost trajektorije



Rezultati u slijedu slika:

- gotovo svi znakovi su pronađeni
- dvije krive detekcije u sekvenci sa 100000 slika
- lokalizacijska pogreška: 17%→12%

Dinamička analiza 3D scena: Cycab 2007

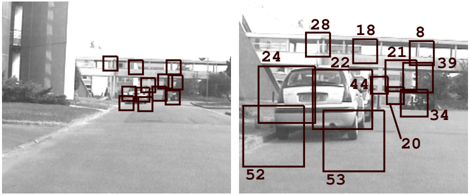
Primjer: sustav za autonomnu navigaciju

- autonomna navigacija u urbanim sredinama vozilom Cycab
- odvojene faze učenja puta i navigacije
- koristi se isključivo kalibrirana perspektivna kamera!
- upravljanje u **domeni slike** (vizualno navođenje)
- rezultati:
  - učenje 1 m/s, navigacija 2 m/s
    (preko 700m uspješne navigacije po javnoj prometnici)
  - stanovita otpornost na ostale sudionike prometa te promjene osvjetljenja
  - duljina scene nije ograničena postupkom mapiranja
    (disk od 1TB dovoljan za 30000 km)

Dinamička analiza 3D scena: Cycab 2007

2D analiza:

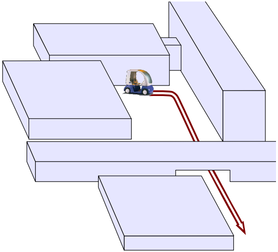
- pronalaženje korespondencija invarijantnim deskriptorima (samo pri inicijalizaciji!)
- pronalaženje značajki (Harris)
- praćenje značajki (KLT)

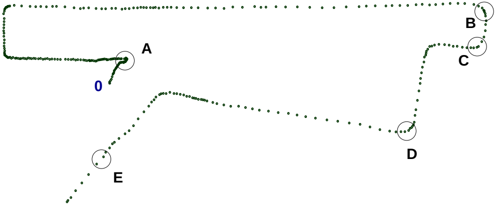


Dinamička analiza 3D scena: Cycab 2007

Interni prikaz scene:

- statičan, dobiven u fazi učenja
- organizacija: topologija lokalnih (metričkih) 3D rekonstrukcija
- metode: procjena relativne orijentacije, estimiranje strukture scene i kretanja kamere





Dinamička analiza 3D scena: Cycab 2007

3D interpretacija: predviđanje 2D položaja značajki koje trenutno ne pratimo na temelju korespondencija sa slikama iz slijeda za učenje

Dinamička analiza 3D scena: ostali primjeri

- stereoskopska odometrija (2013/4)
- povezivanje EuroRAP videa (2011)
- detekcija i raspoznavanje razdjelne linije (2010)
- stabilizacija videa pribavljenog iz zraka (2012)
- stvaranje mozaika ceste (2010)
- estimiranje OD matrica (2012)

Detekcija značajki

Značajka: istaknuti dio slike

- izgled i položaj **prepoznatljivi** iz različitih dijelova scene (kut, blob, ...)
- zbog efikasnije implementacije, najčešći kvadratni oblik (koriste se i pravokutnici, paralelogrami, krugovi, ...)

veličina značajke: kompromis između **otpornosti na zaklanjanje** i neprepoznatljivosti

svojstveno mjerilo i rotacija: noviji koncept, problemi u stvarnom vremenu (GPGPU!)

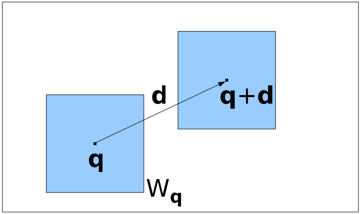
Detekcija značajki

Harrisovi kutevi: unatoč poodmakloj dobi (1988), jedan od najefikasnijih pristupa s fiksnim mjerilom

Ideja: pronaći točke $\mathbf{q}_{zn}$ koje maksimiraju lokalnu prepoznatljivost susjedstva zadane veličine

Promotrimo ``autokorelacijsku funkciju'' (zapravo **SSD**):

$$E_{\mathbf{q}}(\mathbf{d}) = \sum_{x,y \in W(\mathbf{q})} [\mathrm{I}(x + d_x, y + d_y) - \mathrm{I}(x, y)]^2$$



$E_{\mathbf{q}}(\mathbf{d})$: različitost susjedstva točke $\mathbf{q}$, u ovisnosti o pomaku $\mathbf{d}$

Detekcija značajki

Ideja: značajke su u točkama gdje $E_{\mathbf{q}}$ raste u svim smjerovima!

Tražimo kriterij koji će otkriti točke $\mathbf{q}(x_{zn}, y_{zn})$ gdje to vrijedi.

Rezultat: U svakoj točki slike računa se matrica **G**

- **G** sadrži lokalne iznose momenata drugog reda gradijenta slike

$$\mathbf{G}(\mathbf{q}) = \begin{bmatrix} \sum_W \mathrm{I}_x^2 & \sum_W \mathrm{I}_x\mathrm{I}_y \\ \sum_W \mathrm{I}_x\mathrm{I}_y & \sum_W \mathrm{I}_y^2 \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

Harrisov odziv r sada je: $r(\mathbf{q}) = ab - c^2 - k(a+b)^2$

Konačno, značajke su u lokalnim maksimumima odziva:

$\mathbf{q}_{zn} = max_{\mathbf{q}} r(\mathbf{q})$

Praćenje značajki

Cilj: odrediti kretanje zadane značajke kroz slijed slika

- čest kontekst: stvarno vrijeme, 3D rekonstrukcija traži se i efikasnost i preciznost!

Dva pristupa:

- praćenje uparivanjem (matching) --- uparivanje nezavisno pronađenih značajki, primjenom načela bliskosti
  - uparivanje primjenom NCC, SSD, ...
- diferencijalno praćenje (gradijentna optimizacija)
- kombinacija (grubi položaj: uparivanje; fini položaj: diferencijalno praćenje)

Predikcija trenutnog položaja može biti velika pomoć:

- 2D predikcija je OK, ali ne pomaže u drugom okviru!
- agresivniji postupci (SLAM, PTAM) predikciju temelje na 3D modelu gibanja i scene

Praćenje značajki

Diferencijalno praćenje:

- položaj (i oblik) značajke izražavamo parametrima $\mathbf{p}$
  - transformirana značajka: $\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \mathrm{I}(\mathbf{W}(\mathbf{x}; \mathbf{p}))$
  - npr. $\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \mathrm{I}(\mathbf{x} + \mathbf{p})$
- pretpostavka: početni oblik i položaj približno poznati
  - početna aproksimacija: $\mathrm{I}_W(\mathbf{x}, \mathbf{p}_0)$
- tražimo $\mathbf{p}$ koji **minimizira** SSD s referentnim izgledom I_R:

$$\hat{\Delta\mathbf{p}} = \arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} (\mathrm{I}_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) - \mathrm{I}_R(\mathbf{x}))^2$$

  - I_R može pribavljena u prethodnoj slici, prvoj slici, u nekom drugom eksperimentu, ili može biti sintetizirana
- rješenje se temelji na iterativnoj gradijentnoj optimizaciji (Lucas-Kanade, 1981)

Praćenje značajki

Algoritam:

- razvijamo $\mathrm{I}_W(\mathbf{x}, \mathbf{p})$ po Tayloru, u ovisnosti o $\Delta\mathbf{p}$:

$$\mathrm{I}_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) \approx \mathrm{I}_W(\mathbf{x}, \mathbf{p}) + \frac{\partial \mathrm{I}_W}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p}) \cdot \Delta\mathbf{p}$$

- izražavamo skalarni rezidual (funkciju cilja):

$$\begin{aligned} R(\Delta\mathbf{p}) &= \sum_{\mathbf{x}} (\mathrm{I}_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) - \mathrm{I}_R(\mathbf{x}))^2 \\ &= \sum_{\mathbf{x}} \left(\mathrm{I}_W(\mathbf{x}, \mathbf{p}) + \frac{\partial \mathrm{I}_W}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p}) \cdot \Delta\mathbf{p} - \mathrm{I}_R(\mathbf{x}) \right)^2 \end{aligned}$$

- zbog jednostavnosti, uvodimo novu notaciju:

$$R(\Delta\mathbf{p}) = \sum_{\mathbf{x}} (e + \mathbf{g}^\top \Delta\mathbf{p})^2$$

Praćenje značajki

Algoritam (2):

- Tražimo $\Delta\mathbf{p}$ koji **minimizira** rezidual kvadrirane L2 norme:

$$\Delta\hat{\mathbf{p}} = min_{\Delta\mathbf{p}} R(\Delta\mathbf{p}) = min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} (e + \mathbf{g}^\top \Delta\mathbf{p})^2$$

- minimum dobivamo izjednačavanjem derivacije s nulom:

$$\frac{\partial R(\Delta\hat{\mathbf{p}})}{\partial \Delta\hat{\mathbf{p}}} = \sum_{\mathbf{x}} 2 \cdot (e + \mathbf{g}^\top \Delta\hat{\mathbf{p}}) \cdot \mathbf{g}^\top = \mathbf{0}^\top$$

- rezultat iteracije (optimalni $\Delta\mathbf{p}$) rješenje linearnog sustava:

$$\sum_{\mathbf{x}} \mathbf{g}e + \sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top \Delta\hat{\mathbf{p}} = \mathbf{0}$$

Praćenje značajki

Algoritam (3):

- Iteriranje se prekida kad:
  - ili poboljšanje $\|\Delta\hat{\mathbf{p}}\|$ postane beznačajno,
  - ili novi položaj bude izvan slike,
  - ili matrica $\sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top$ postane slabo kondicionirana (Harris!)
- postupak je općenit, može se primijeniti za razne transformacije:
  - translacija (2DOF): $\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \mathrm{I}(\mathbf{x} + \mathbf{p})$
  - geometrijska afina transformacija (6DOF):
    $\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \mathrm{I}(\mathbf{A}(\mathbf{p}) \cdot \mathbf{x} + \mathbf{d}(\mathbf{p}))$
  - fotometrijska afina transformacija (2DOF): $\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \lambda\mathrm{I}(\mathbf{x}) + \delta$

Praćenje značajki

Implementacija (Shi-Tomasi '94):

- prvo izvesti translatorno praćenje u odnosu na pojavljivanje značajke u prethodnoj slici: $\mathrm{I}_R^1(\mathbf{x}) = \mathrm{I}_{i-1}(\mathbf{x}, \mathbf{p}_{i-1})$, $\mathbf{p} = (x, y)$
- zatim poboljšati rezultat afinim praćenjem u odnosu na izgled *pri detekciji*: $\mathrm{I}_R^2(\mathbf{x}) = \mathrm{I}_0(\mathbf{x}, \mathbf{p}_0)$, $\mathbf{p} = (x, y, \phi, m, ...)$
- odustati od praćenja ako $\|R(\Delta\mathbf{p})\| > r_{TH}$

Praćenje značajki

Implementacija za slike pribavljene iz vozila u pokretu (1)

- pretpostavka: perspektivne slike, kretanje unaprijed (afinih izobličenja i rotacije uglavnom **nema**)
- transformacija s 5 SS:

$$\mathrm{I}_W(\mathbf{x}, \mathbf{p}) = \lambda \cdot \mathrm{I}(m \cdot \mathbf{x} + \mathbf{d}) + \delta = \mathrm{U}(\mathrm{I}(\mathbf{T}(\mathbf{x}, \mathbf{q})), \mathbf{r})$$

- $\mathbf{p} = (\mathbf{q}, \mathbf{r})$, $\mathbf{q} = (m, \mathbf{d})$, i $\mathbf{r} = (\lambda, \delta)$.
- $\mathrm{U}(\mathrm{I}_T, [\lambda, \delta]^\top) = \lambda \cdot \mathrm{I}_T + \delta$ (I_T ... siva razina piksela)
- $\mathbf{T}(\mathbf{x}, [m, \mathbf{d}^\top]^\top) = m \cdot \mathbf{x} + \mathbf{d}$ (**x** ... položaj piksela)

Praćenje značajki

Implementacija za slike pribavljene iz vozila u pokretu (2)

- transformacija s 5 SS:

$$\mathrm{I}_W(\mathbf{x},\mathbf{p}) = \lambda \cdot \mathrm{I}(m \cdot \mathbf{x} + \mathbf{d}) + \delta = \mathrm{U}(\mathrm{I}(\mathbf{T}(\mathbf{x},\mathbf{q})),\mathbf{r})$$

- za generalnu formulaciju moramo odrediti $\frac{\partial \mathrm{I}_W}{\partial \mathbf{p}} = [\frac{\partial \mathrm{U}}{\partial \mathbf{q}} \frac{\partial \mathrm{U}}{\partial \mathbf{r}}]$
- određujemo drugi član uz pokrate $\mathrm{I}_T = \mathrm{I}(\mathbf{x_T})$, $\mathbf{x_T} := \mathbf{T}(\mathbf{x},\mathbf{q})$:
  $\frac{\partial \mathrm{U}}{\partial \mathbf{r}}(\mathrm{I}_T,\mathbf{r}) = \left[\begin{array}{cc} \mathrm{I}_T & 1 \end{array}\right]$
- prvi član je malo zahtjevniji (uvodimo $\mathrm{I}_T^{\mathbf{x}} = \frac{\partial \mathrm{I}}{\partial \mathbf{x}}(\mathbf{x_T})$)

$$\begin{aligned}\frac{\partial \mathrm{U}}{\partial \mathbf{q}}(\mathrm{I}_T,\mathbf{r}) &= \frac{\partial \mathrm{U}}{\partial \mathrm{I}}(\mathrm{I}_T,\mathbf{r}) \cdot \frac{\partial \mathrm{I}}{\partial \mathbf{T}}(\mathbf{x_T}) \cdot \frac{\partial \mathbf{T}}{\partial \mathbf{q}}(\mathbf{x},\mathbf{q}) \\ &= \lambda \cdot \mathrm{I}_T^{\mathbf{x}} \cdot \left[\begin{array}{ccc} x_1 & 1 & 0 \\ x_2 & 0 & 1 \end{array}\right] = \lambda \left[\begin{array}{ccc} \mathrm{I}_T^{\mathbf{x}}\mathbf{x} & \mathrm{I}_T^{x1} & \mathrm{I}_T^{x2} \end{array}\right]\end{aligned}$$

Praćenje značajki

Implementacija za slike pribavljene iz vozila u pokretu (3):

- Konačno, dobivamo izraz za $\mathbf{g}$:

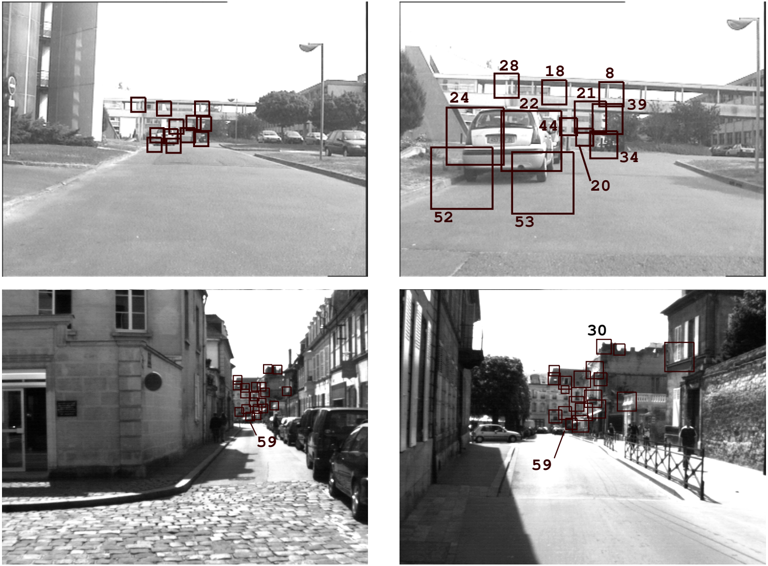
$$\mathbf{g}^\top = \left[\begin{array}{ccccc} \lambda \mathrm{I}_T^{\mathbf{x}}\mathbf{x} & \lambda \mathrm{I}_T^{x1} & \lambda \mathrm{I}_T^{x2} & \mathrm{I}_T & 1 \end{array}\right]$$

- Dobiveni izraz koristimo u općenitom iterativnom postupku:

$$\sum_{\mathbf{x}} \mathbf{g} e + \sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top \Delta\hat{\mathbf{p}} = \mathbf{0}$$

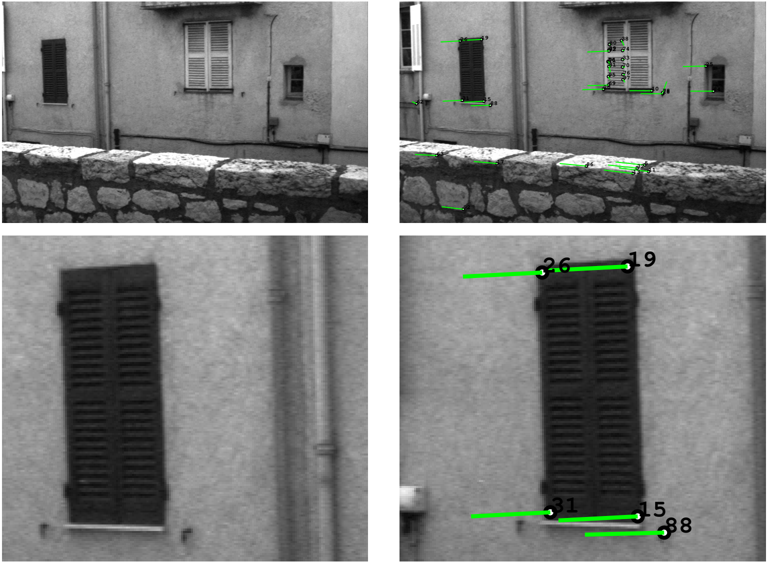
Praćenje značajki

Rezultati: pouzdano praćenje u dugim nizovima slika



Praćenje značajki

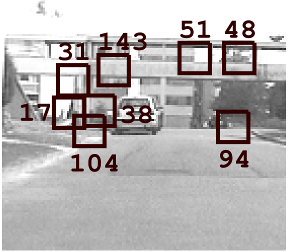
Rezultati: tolerancija velikih pomaka (višerezolucijska obrada)



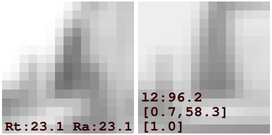
Praćenje značajki

Problemi:

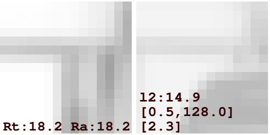
- odlazak u krivi lokalni minimum uslijed akumulirane pogreške
- značajka može ``preskočiti'' na neki drugi dio scene uslijed:
  - zaklanjanja praćenog dijela scene
  - neprikladne 3D strukture praćenog dijela scene



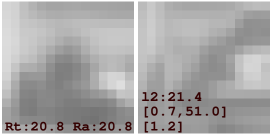
tony4:1−220



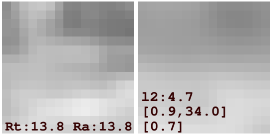
#17@tony4:190



#31@tony4:205



#38@tony4:55



#104@tony4:56

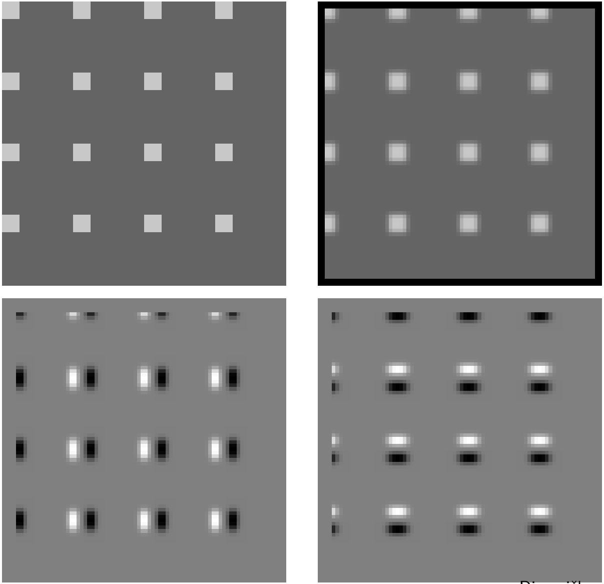
Praćenje značajki

Diskusija:

- glavni doprinos diferencijalnog praćenja: podpikselska preciznost
  - preliminarni rezultati na KITTI-ju: funkcija cilja vizualne odometrije 25%$\downarrow$
- postupci temeljeni na uparivanju vjerojatno nude bolju performansu na razini piksela
- zanimljivo rješenje: hibridni postupci

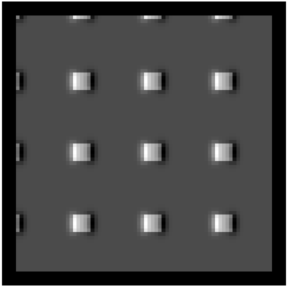
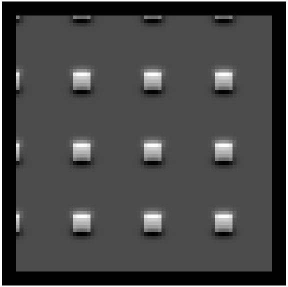
Praćenje značajki: demonstracija

Originalna slika, zaglađena slika i komponente gradijenta:



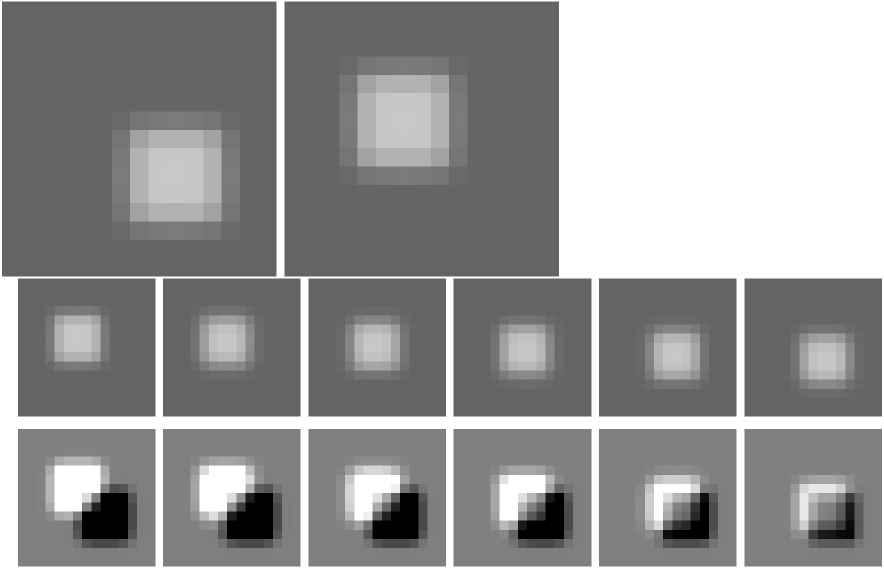
Praćenje značajki: demonstracija (2)

Provjera poravnavanja derivacija s originalnom slikom:

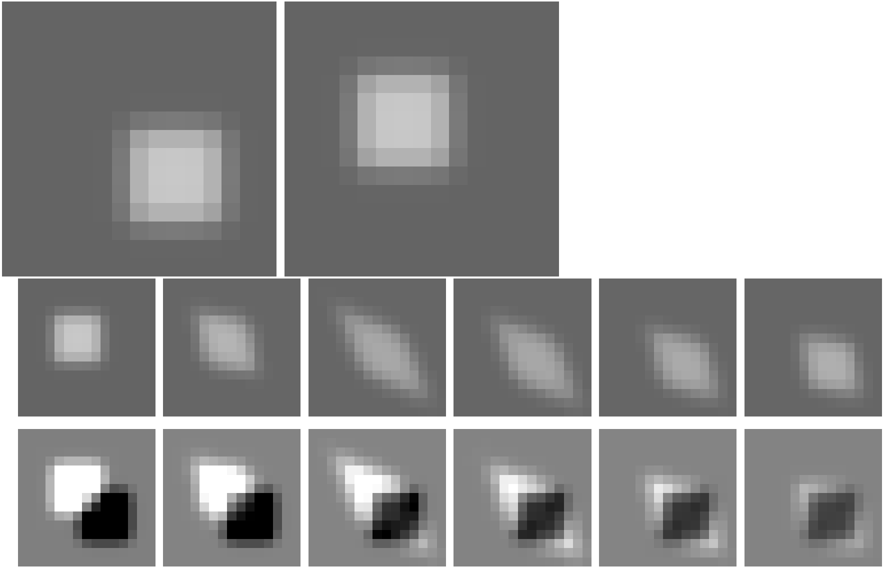
Praćenje značajki: demonstracija (3)

Postupak konvergencije za model s 2DOF:



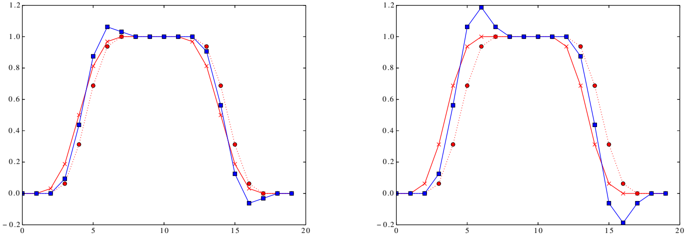
Praćenje značajki: demonstracija (4)

Postupak konvergencije za model s 8DOF:



Praćenje značajki: demonstracija (5)

Točnost aproksimacije posmaka Taylorovim razvojem prvog reda:

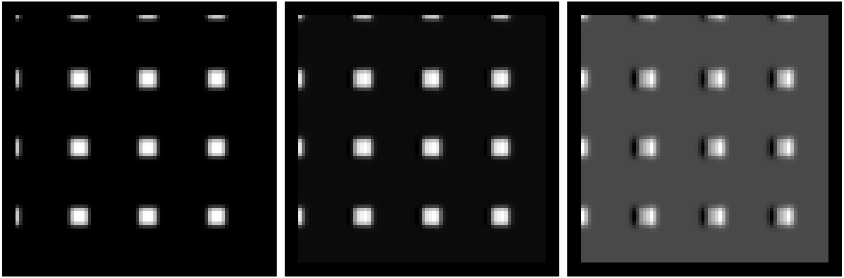


Grafovi prikazuju originalni signal (crtkano), posmaknuti signal (crveno), i aproksimaciju prvog reda (plavo).

Posmak iznosi 0.5 piksela (lijevo) odnosno 1 piksel (desno).

Praćenje značajki: demonstracija (6)

Točnost aproksimacije posmaka Taylorovim razvojem prvog reda:



Posmak iznosi 0 piksela (lijevo), 1 piksel (sredina) odnosno 5 piksela (desno).

Optimizacija: KLT

Podsjetimo se na jednadžbu pomaka algoritma KLT:

$$\sum_{\mathbf{x}} \mathbf{g} e + \sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top \Delta\hat{\mathbf{p}} = \mathbf{0}$$

Gradijent $\mathbf{g}$ i greška e odgovaraju sljedećim izrazima:

$$\mathbf{g}^\top = \frac{\partial \mathrm{I}_W}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p})$$

$$e = \mathrm{I}_W(\mathbf{x}, \mathbf{p}) - \mathrm{I}_R(\mathbf{x})$$

Odgovarajući pomak traži stacionarnu točku kriterija:

$$\Delta\hat{\mathbf{p}} = -\left(\sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top\right)^{-1} \sum_{\mathbf{x}} \mathbf{g} e$$

Optimizacija: gradijentni spust

Podsjetimo se na naš optimizacijski kriterij:

$$R(\mathbf{p}) = \sum_{\mathbf{x}} \left(\mathrm{I}_W(\mathbf{x}, \mathbf{p}) - \mathrm{I}_R(\mathbf{x}) \right)^2$$

Nađimo gradijent s obzirom na parametre:

$$\frac{\partial \mathrm{R}}{\partial \mathbf{p}} = 2 \sum_{\mathbf{x}} (\mathrm{I}_W(\mathbf{x}, \mathbf{p}) - \mathrm{I}_R(\mathbf{x})) \cdot \frac{\partial \mathrm{I}_W}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p})$$

Biramo pomak koji smanjuje kriterij:

$$\Delta\hat{\mathbf{p}} = -\delta \cdot \frac{\partial \mathrm{R}}{\partial \mathbf{p}} = -\delta \cdot \sum_{\mathbf{x}} \mathbf{g} e$$

Optimizacija: usporedba

Pomak prema algoritmu KLT (Gauss-Newton):

$$\Delta\hat{\mathbf{p}}^{KLT} = -\left(\sum_{\mathbf{x}} \mathbf{g}\mathbf{g}^\top\right)^{-1} \sum_{\mathbf{x}} \mathbf{g}e$$

Pomak prema gradijentnom spustu:

$$\Delta\hat{\mathbf{p}}^{GD} = -\delta \cdot \sum_{\mathbf{x}} \mathbf{g}e$$

Hvala na pažnji!

Pitanja?