

# 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:

- računaska 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:

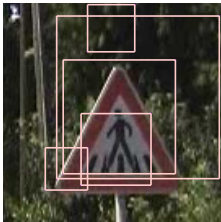
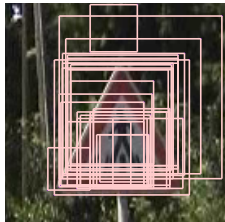
- 2D analiza (pronalaženje, uparivanje, praćenje, raspoznavanje)
- interni prikaz scene (statičan ili dinamičan)
- 3D interpretacija (kretanje kamere, struktura scene)

# 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

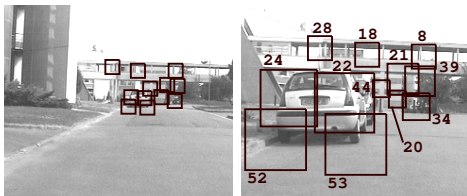
**Primjer:** sustav za autonomnu navigaciju

- 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:

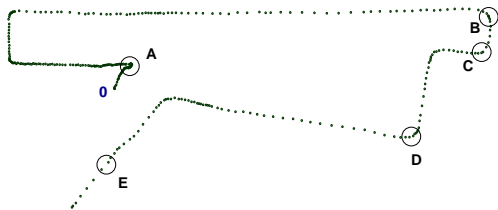
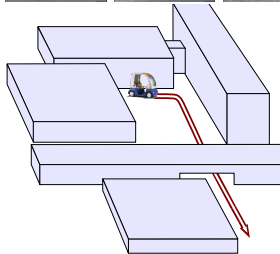
- pronalaženje korespondencija invarijantnim deskriptorima (samo pri inicijalizaciji!)
- pronalaženje značajki (Harris)
- praćenje značajki (KLT)



# 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!)

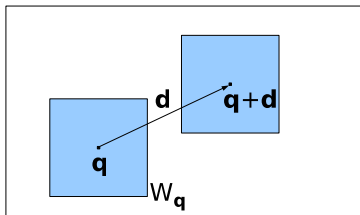
## DETEKCIJA ZNAČAJKI

**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**):

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



$E_{\mathbf{q}}(\mathbf{d})$ : 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  $\mathbf{G}$

- $\mathbf{G}$  sadrži lokalne iznose momenata drugog reda gradijenta slike

$$\mathbf{G}(\mathbf{q}) = \begin{bmatrix} \sum_W I_x^2 & \sum_W I_x I_y \\ \sum_W I_x I_y & \sum_W I_y^2 \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \quad (1)$$

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

$$\hat{\Delta\mathbf{p}} = \arg \min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} (I_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) - I_R(\mathbf{x}))^2$$

- $I_R$  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  $I_W(\mathbf{x}, \mathbf{p})$  po Tayloru, u ovisnosti o  $\Delta \mathbf{p}$ :

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

- izražavamo skalarni rezidual (funkciju cilja):

$$\begin{aligned} R(\Delta \mathbf{p}) &= \sum_{\mathbf{x}} (I_W(\mathbf{x}, \mathbf{p} + \Delta \mathbf{p}) - I_R(\mathbf{x}))^2 \\ &= \sum_{\mathbf{x}} (I_W(\mathbf{x}, \mathbf{p}) + \frac{\partial I_W}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p}) \cdot \Delta \mathbf{p} - I_R(\mathbf{x}))^2 \end{aligned}$$

- zbog jednostavnosti, uvodimo novu notaciju:

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

## PRAĆENJE ZNAČAJKI

### Algoritam (2):

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

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

- minimum dobivamo izjednačavanjem derivacije s nulom:

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

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

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



# PRAĆENJE ZNAČAJKI

## Algoritam (3):

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

# PRAĆENJE ZNAČAJKI

## Implementacija (Shi-Tomasi '94):

- prvo izvesti translatorno praćenje u odnosu na pojavljivanje značajke u prethodnoj slici:  $I_R^1(\mathbf{x}) = I_{i-1}(\mathbf{x}, \mathbf{p}_{i-1})$ ,  $\mathbf{p} = (x, y)$
- zatim poboljšati rezultat afinim praćenjem u odnosu na izgled *pri detekciji*:  $I_R^2(\mathbf{x}) = I_0(\mathbf{x}, \mathbf{p}_0)$ ,  $\mathbf{p} = (x, y, \phi, m, \dots)$
- 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:

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

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

# PRAĆENJE ZNAČAJKI

**Implementacija** za slike pribavljene iz vozila u pokretu (2)

- transformacija s 5 SS:

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

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

$$\begin{aligned} \frac{\partial U}{\partial \mathbf{q}}(I_T, \mathbf{r}) &= \frac{\partial U}{\partial I}(I_T, \mathbf{r}) \cdot \frac{\partial I}{\partial \mathbf{T}}(\mathbf{x}_T) \cdot \frac{\partial \mathbf{T}}{\partial \mathbf{q}}(\mathbf{x}, \mathbf{q}) \\ &= \lambda \cdot I_T^{\mathbf{x}} \cdot \begin{bmatrix} x_1 & 1 & 0 \\ x_2 & 0 & 1 \end{bmatrix} = \lambda \begin{bmatrix} I_T^{\mathbf{x}} \mathbf{x} & I_T^{\mathbf{x}1} & I_T^{\mathbf{x}2} \end{bmatrix} \end{aligned}$$

## PRAĆENJE ZNAČAJKI

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

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

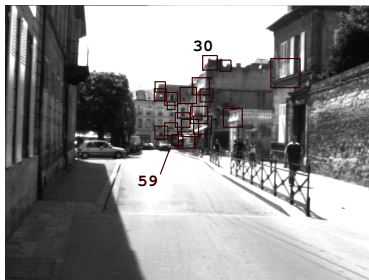
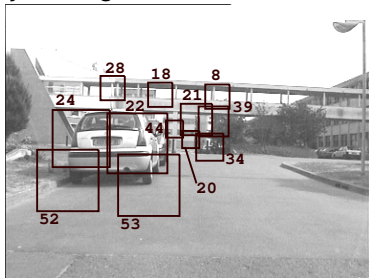
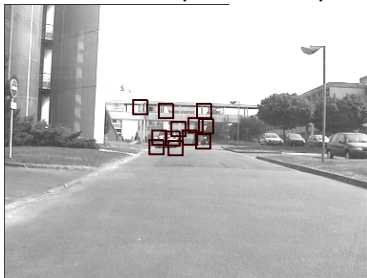
$$\mathbf{g}^T = \left[ \lambda \mathbf{I}_T^{\mathbf{x} \times \mathbf{x}} \quad \lambda \mathbf{I}_T^{\mathbf{x} \times 1} \quad \lambda \mathbf{I}_T^{\mathbf{x} \times 2} \quad \mathbf{I}_T \quad 1 \right] \quad (2)$$

- Dobiveni izraz koristimo u općenitom iterativnom postupku:

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

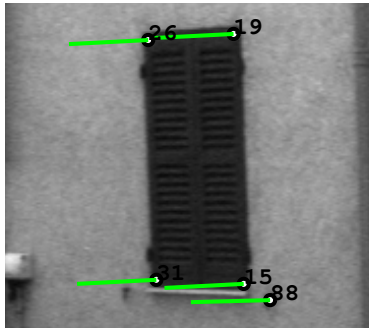
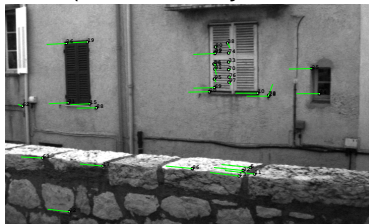
# PRAĆENJE ZNAČAJKI

**Rezultati:** pouzdano praćenje u dugim nizovima slika



# PRAĆENJE ZNAČAJKI

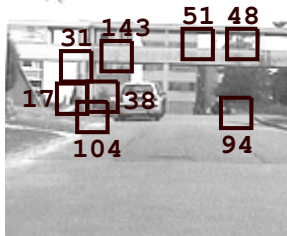
**Rezultati:** tolerancija velikih pomaka (višerezolucijska obrada)



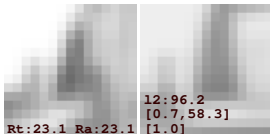
# PRAĆENJE ZNAČAJKI

## Problemi:

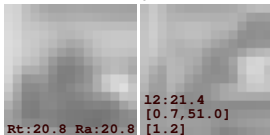
- odlazak u krivi lokalni minimum uslijed akumulirane pogreške
- značajka može "preskočiti" na neki drugi dio scene uslijed:
  - zaklanjanja praćenog dijela scene
  - neprikladne 3D strukture praćenog dijela scene



tony4:1-220



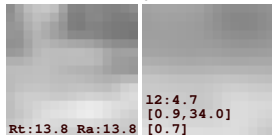
#17@tony4:190



#38@tony4:55



#31@tony4:205



#104@tony4:56



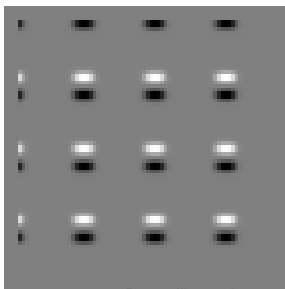
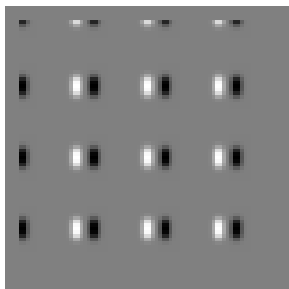
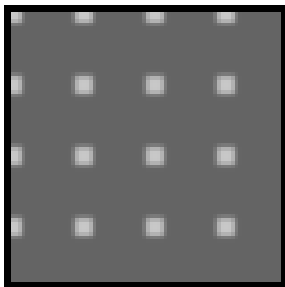
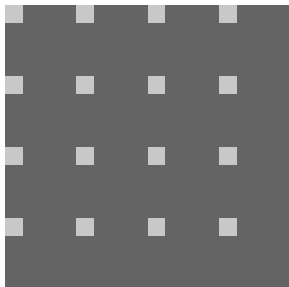
# PRAĆENJE ZNAČAJKI

## Diskusija:

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

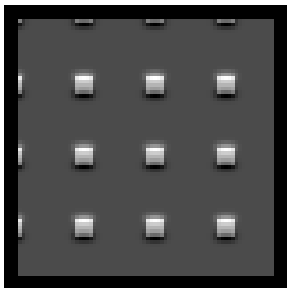
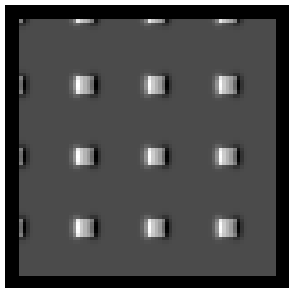
# PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Originalna slika, zaglađena slika i komponente gradijenta:



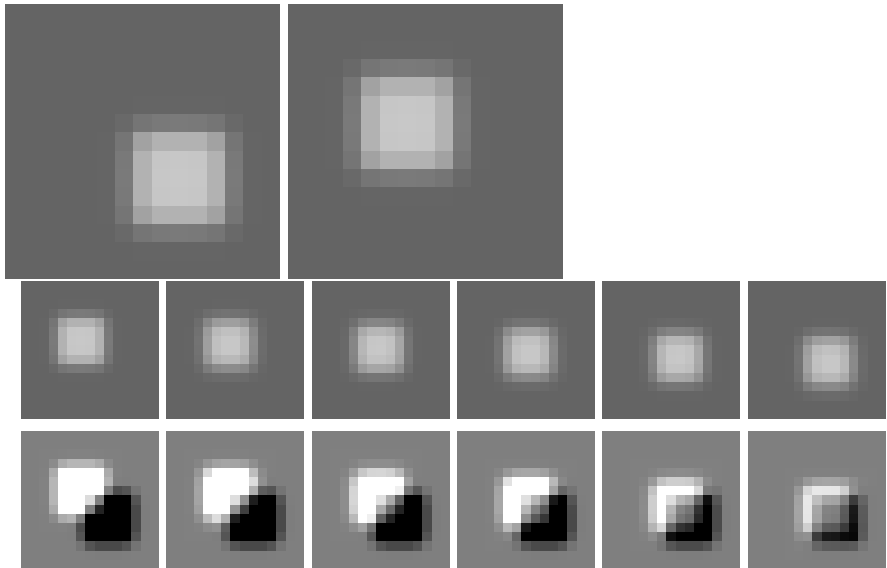
## PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Provjera poravnavanja derivacija s originalnom slikom:



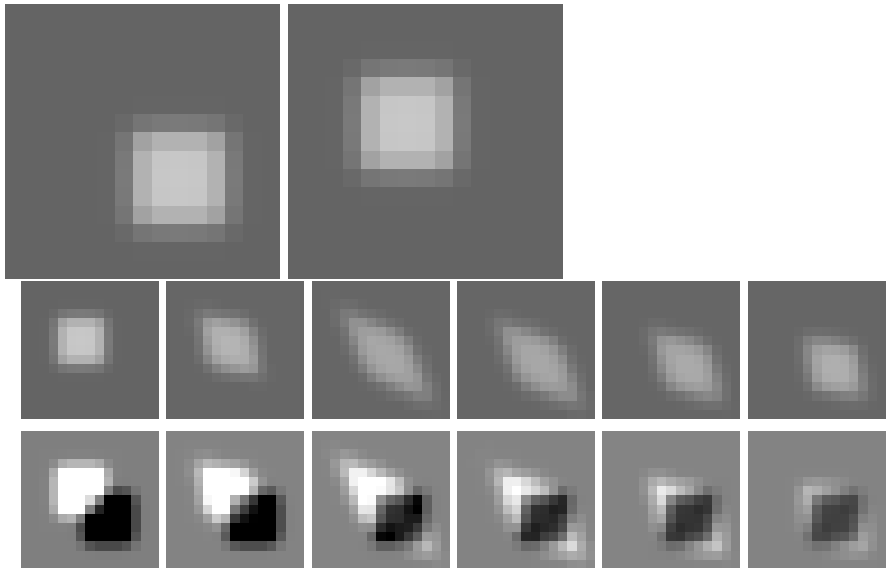
# PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Postupak konvergencije za model s 2DOF:



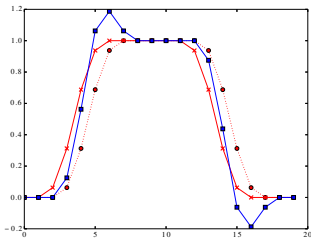
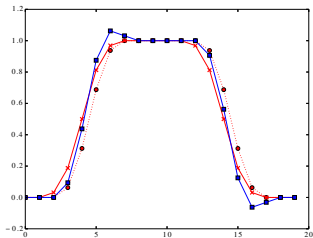
# PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Postupak konvergencije za model s 8DOF:



# PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Točnost aproksimacije posmaka Taylorovim razvojem prvog reda:

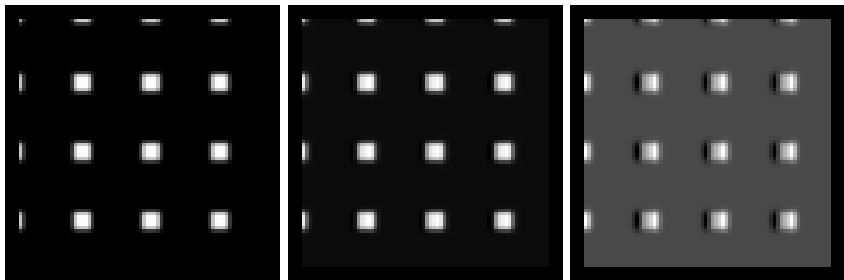


Grafovi prikazuju originalni signal (crtkano), posmaknuti signal (crveno), i aproksimaciju prvog reda (plavo).

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

## PRAĆENJE ZNAČAJKI, IMPLEMENTACIJA

Točnost aproksimacije posmaka Taylorovim razvojem prvog reda:

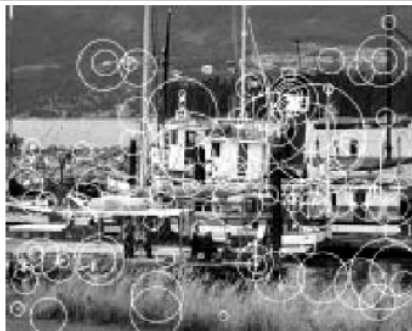


Posmak iznosi 0 piksela (lijevo), 1 piksel (sredina) odnosno 5 piksela (desno).

# INVARIJANTNOST NA MJERILO

**Cilj:** Odrediti korespondenciju međusobno razmaknutih pogleda (wide-baseline matching)

- ▣ različiti smjerovi gledanja (oblik!)
- ▣ različita udaljenost kamera od scene (mjerilo!)



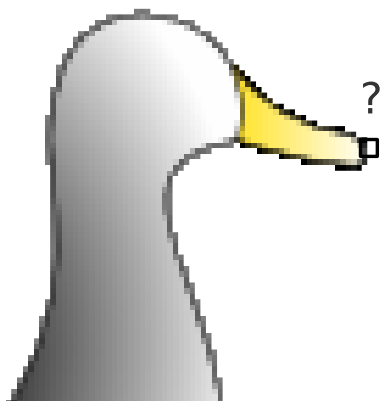
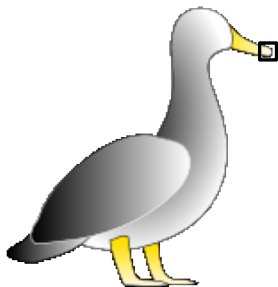
(a) Scale change of 3.9 and rotation of  $17^\circ$ .



# INVARIJANTNOST NA MJERILO

## Što fali Harrisu?:

- vidjeli smo da je odziv rotacijski invarijantan, ...
- ali ne i invarijantan na mjerilo!



## INVARIJANTNOST NA MJERILO

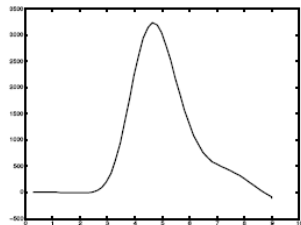
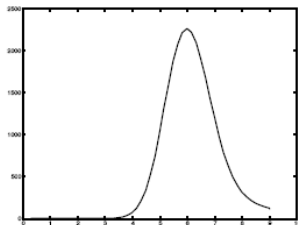
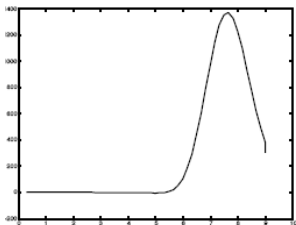
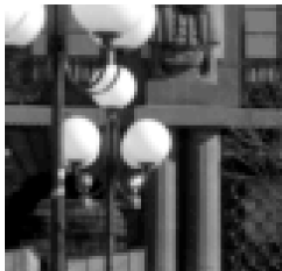
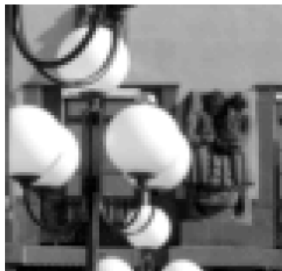
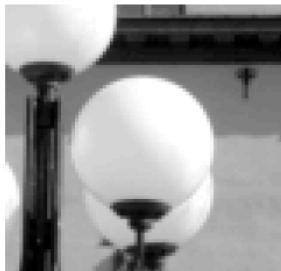
Određivanje korespondencija preko jako razmaknutih pogleda *zahtijeva* analizu slika u **prostoru mjerila** (scale-space)!

- u svakoj izlučenoj značajci moramo procijeniti svojstveno mjerilo, kako bismo mogli testirati korespondenciju na međusobno usklađenim mjerilima
- to je čvrsti zahtjev, ne samo optimizacija.



# INVARIJANTNOST NA MJERILO

Ideja: svojstveno mjerilo u središtu slike kao ekstrem odziva  $LoG$



# INVARIJANTNOST NA MJERILO

## Analogija:

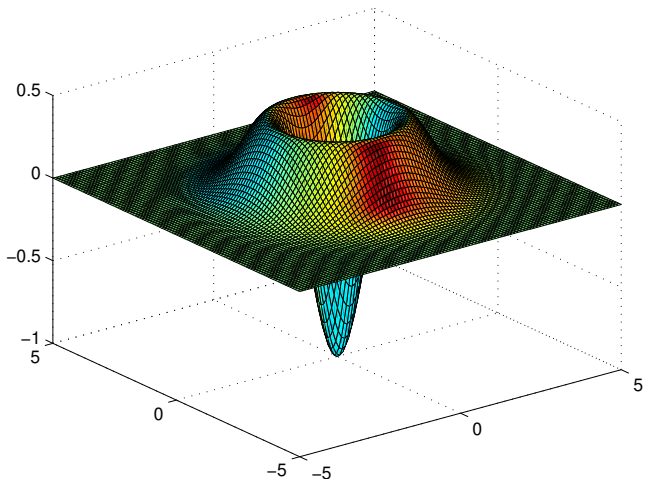
(zaglađena) slika	potencijal
gradijent slike	polje
divergencija gradijenta	gustoća naboja

Svojstveno mjerilo  $\sigma$ :

- mjerilo kod kojeg se promatrano susjedstvo može objasniti najvećom gustoćom naboja u središtu
- koliko susjedstvo treba smanjiti da postane što sličnije izdvojenom pikselu na kontrastnoj podlozi?

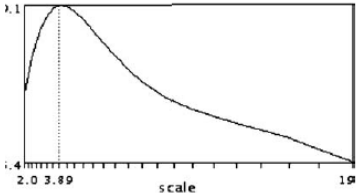
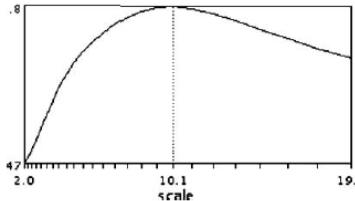
# INVARIJANTNOST NA MJERILO

**Zašto to radi?:** LoG daje ekstremni odziv na blobu prilagođene veličine



# INVARIJANTNOST NA MJERILO

Još jedan primjer:



# INVARIJANTNOST NA MJERILO

**SIFT** deskriptor: konkatencija histograma gradijenta s 8 pretinaca

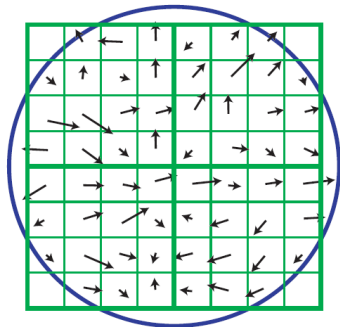
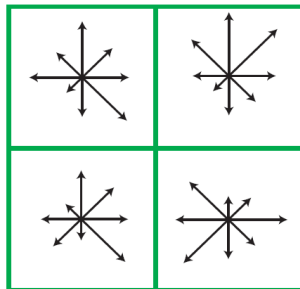


Image gradients



Keypoint descriptor

[Lowe2004]

## INVARIJANTNOST NA MJERILO

Veličina i orijentacija kvadratnog polja ovise o svojstvenom mjerilu i orijentaciji ključne točke:



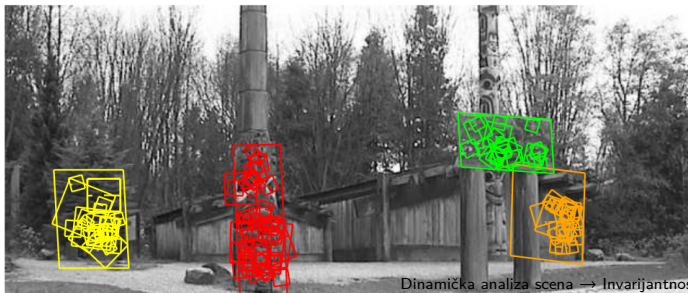
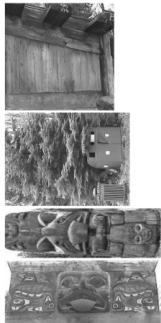


# INVARIJANTNOST NA MJERILO

## Okvirni postupak pronalaženja korespondencija:

1. pronaći istaknute točke (keypoint extraction)
2. dodijeliti mjerilo, orijentaciju, izračunati **deskriptore**
3. odrediti korespondencije u  $O(n^2)$

# INVARIJANTNOST NA MJERILO

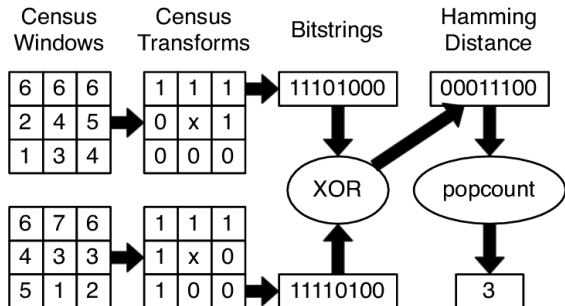


Dinamička analiza scena → Invarijantnost na mjerilo (10) 46/??

# ROBUSNO UPARIVANJE

## Census:

1. lokalnu okolinu predstaviti binarnim deskriptorom
  - bit je 1 ako je piksel veći od središnjeg (inače 0)
2. sličnost deskriptora modeliramo Hammingovom udaljenošću
3. posebno prikladno u stvarnim scenama u kombinaciji sa SGM-om
  - može doći do velike razlike prosječne sive razine u korespondentnim regijama



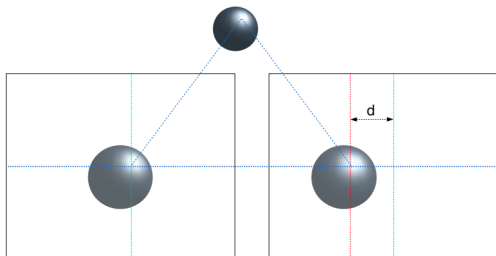
## GLOBALNI KRITERIJ

3D svijet nije proizvoljan, ne izgleda kao Dirichletova funkcija

1. Ideja: ohrabriti glatke rekonstrukcije
2. npr. kažnjavati razlike u disparitetu između susjednih piksela
3. ne smijemo pretjerati jer je zaklanjanje česta pojava

Ovu ideju ćemo posebno razmotriti u kontekstu rektificiranog sterea

- znamo da se korespondencija nalazi u istom retku slike
- u svakom pikselu tražimo disparitet  $d$  koji ovisi o dubini  $z$ .



## GLOBALNI KRITERIJ: FORMULACIJA

Ulaz: gusti volumen korespondencijske cijene  $C$  ( $C$  definirana u svakom pikselu i disparitetu)

Globalna optimizacija: tražimo polje dispariteta  $D$  koje minimizira i) korespondencijsku cijenu, ii) regularizacijski član.

$$E(D) = \sum_{\mathbf{p}} (C(\mathbf{p}, D_{\mathbf{p}})) + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_1 [ [|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] ] + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_2 [ [|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] ]$$

- $C$  - lokalna cijena (npr. L2, L1, Census)
- $D$  - gusto polje dispariteta (predmet minimizacije)
- $P_1, P_2$  hiperparametri,  $P_1 \leq P_2$

Problem: cijena u pikselu  $\mathbf{p}$  ovisi o cijeni u susjedima (koja opet ovisi o cijeni u njihovim susjedima) ---  $O(N_d^{HW})$

## GLOBALNI KRITERIJ: POLUGLOBALNO UPARIVANJE (1)

Ideja: provesti kompromis s obzirom na pristup utemeljen na gruboj sili

- 1D optimizacija dinamičkim programiranjem

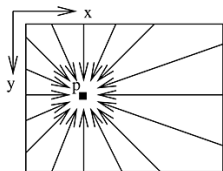
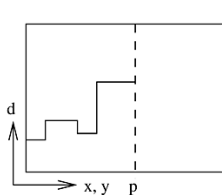
Algoritam se odvija u nekoliko koraka:

1. izračun lokalne cijene korespondencije  $C$  (npr. L2, L1, Census)
  - izlaz: volumen cijene  $H \times W \times N_d$
2. agregacija cijene uzduž  $R$  puteva ( $R=8, R=16$ )
  - oblik dinamičkog programiranja (srodno Viterbijevom algoritmu)
3. odabir dispariteta prema najboljoj globalnoj cijeni.

Ključni korak 2 ćemo pobliže razmotriti.

## GLOBALNI KRITERIJ: POLUGLOBALNO UPARIVANJE (3)

Rješenje provodimo dinamičkim programiranjem u 8 (ili 16) smjerova:



[hirschmueller05cvpr]

Ideja: gusto polje  $L_r$  prati minimalnu **ukupnu** cijenu za svaki piksel  $\mathbf{p}$  i disparitet  $d$  s obzirom na sve prethodne piksele u smjeru  $r$ :

$$L_r(\mathbf{p}, d) = C(\mathbf{p}, d) + \min(L_r(\mathbf{p} - \mathbf{r}, d),$$

$$L_r(\mathbf{p} - \mathbf{r}, d - 1) + P_1,$$

$$L_r(\mathbf{p} - \mathbf{r}, d + 1) + P_1,$$

$$\min_i L_r(\mathbf{p} - \mathbf{r}, i) + P_2) - \min_k L_r(\mathbf{p} - \mathbf{r}, k)$$

## GLOBALNI KRITERIJ: POLUGLOBALNO UPARIVANJE (4)

Algoritam agregacije:

1. Izračunati  $L_r$  za svaki od  $r$  smjerova
2. Izračunati volumen globalne cijene  $S(\mathbf{p}, d) = \sum_r L_r(\mathbf{p}, d)$

Odabir dispariteta:

- $D_{\mathbf{p}} = \arg \min_d S(\mathbf{p}, d)$
- poboljšanje 1: agregirati  $S$  za oba redoslijeda ulaznih slika
- poboljšanje 2: odrediti podpikselski disparitet kao argmin polinoma kroz susjedne disparitete

Vremenska složenost:  $O(HWN_d)$

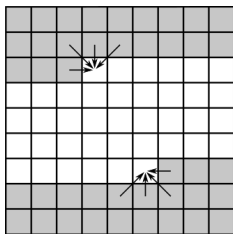
- sklopovske implementacije (FPGA) mogu raditi u stvarnom vremenu



## GLOBALNI KRITERIJ: POLUGLOBALNO UPARIVANJE (3)

Optimirana izvedba izbjegava potrebu za pohranjivanjem  $L_r$ -ova

Umjesto toga, globalna cijena se agregira izravno u S kroz dva prolaza:



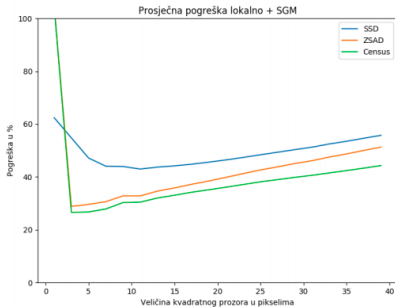
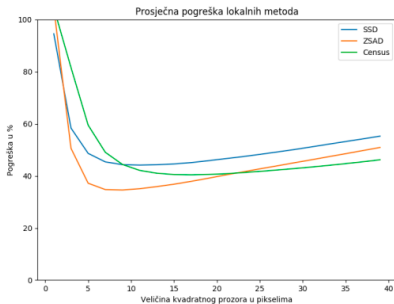
[wikipedia]

Memorijska složenost za dvoprolaznu implementaciju s  $r=8$ :

- Volumeni lokalne i globalne cijene  $C$  i  $S$ :  $2 \times H \times W \times N_d$
- Cacheiranje globalnih cijena kroz tri retka:  $3 \times W \times N_d$
- Za  $W=2000$ ,  $H=1000$ ,  $N_d = 200$ : oko 3GB

# GLOBALNI KRITERIJ: POLUGLOBALNO UPARIVANJE (4)

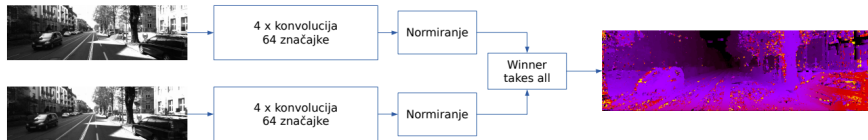
Eksperimentalno vrednovanje potvrđuje vrijednost SGM-a:



[bunjevac17bs]

# NAUČENA KORESPONDENCIJA

Ideja: naučiti konvolucijski model koji svaki piksel predstavlja 64-D opisnikom koji optimira korespondenciju



[orsic17ms]

Kažemo da ovakvi modeli provode korespondencijsko (ili metričko) ugrađivanje

- slični pikseli dobivaju slične opisnike (u smislu skalarnog produkta)

## NAUČENA KORESPONDENCIJA

Ako su dostupni točni dispariteti (KITTI) korespondencijsko ugrađivanje možemo učiti na trojkama okana:

- uzimamo svako okno lijeve slike  $R = I_L(x, y)$
- pronalazimo odgovarajuće okno  $P$  na točnom disparitetu:  
 $P = I_R(x - d, y)$
- pronalazimo slučajno okno  $Q$  na odklonu  $o$  od točnog dispariteta:  $Q = I_R(x - d - o, y)$
- trojka  $\langle R, P, Q \rangle$  čini jedan primjer za učenje

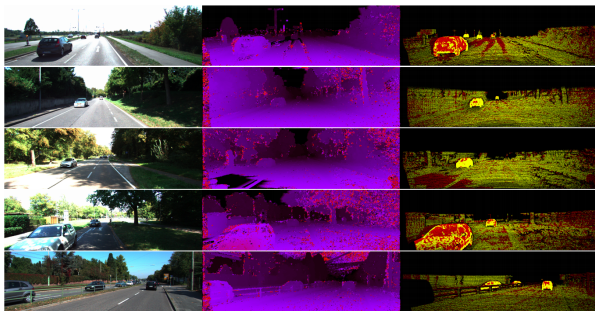
Optimiziramo trojni (engl. triplet) zglobni (engl. hinge) gubitak (bs=128):

$$L(R, P, Q) = \max(0, m + h(R|\theta) \cdot h(Q|\theta) - h(R|\theta) \cdot h(P|\theta))$$

# NAUČENA KORESPONDENCIJA

Model	Točnost - treniranje	Točnost - testiranje
Sive ulazne slike	84.99%	82.32%
Sive ulazne slike - BN	74.51%	71.61%
Ulazne slike u boji	<b>85.50%</b>	<b>82.84%</b>
Ulazne slike u boji - BN	74.40%	71.33%

[orsic17ms]



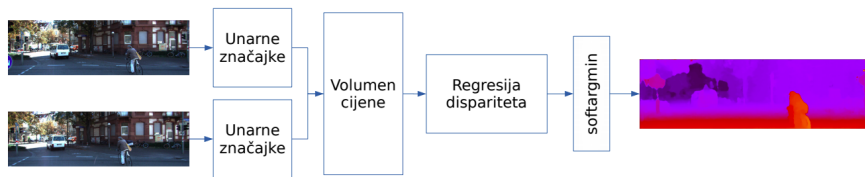
[orsic17ms]

Dobra generalizacija: slučajno inicijalizirani model bolji od SSD-a.

Nakon primjene SGM-a točnost može porasti do 95%.

# NAUČENA KORESPONDENCIJA

Ideja: naučiti konvolucijski model koji integrira glaćenje [kendall17iccv]:



[orsic17ms]

Glađenje zahtijeva 3D konvoluciju  $(x,y,d)$  nad 4D tenzorima

Konačni volumen cijene  $C$  se vodi na soft arg min u svakom pikselu:

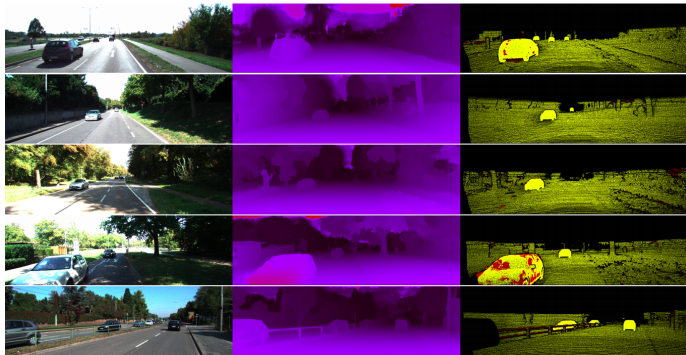
$$d_{xy} = \text{soft arg min}(C_{xy}) = \sum_{d=0}^{D_{\max}} d \cdot \text{softmax}(-C_{xyd})$$

Model ima ogroman broj parametara i velike memorijske zahtjeve: učenje na isječcima  $256 \times 512$ , veličina grupe 1 ili 2.

Samo 200 označenih slika (180: train, 20: val)  $\Rightarrow$  unakrsna validacija.

# NAUČENA KORESPONDENCIJA

Model	Točnost - treniranje	Točnost - testiranje	
Osnovni model	<b>97.98%</b>	<b>97.23%</b>	
Razlika unarnih značajki	97.62%	97.17%	
Skalarni umnožak unarnih značajki	97.95%	97.07%	
Osnovni model - predtreniranje	<b>98.06%</b>	<b>97.31%</b>	
Razlika unarnih značajki - predtreniranje	97.94	97.18	[orsic17ms]



Dobri rezultati na KITTI-ju, ali loša generalizacija.

## Robusni pristupi u računalnom vidu:

1. Houghova transformacija
2. Evaluacija hipoteza dobivenih nad minimalnim slučajnim uzorkom  
Monte Carlo analiza, RANSAC, MLESAC, \*\*\*SAC, LMedSqr
3. Iterativno poboljšanje korištenjem robusnih normi:  
IRLS, M-estimacija

(I to je sve...)



Hvala na pažnji!

Pitanja?