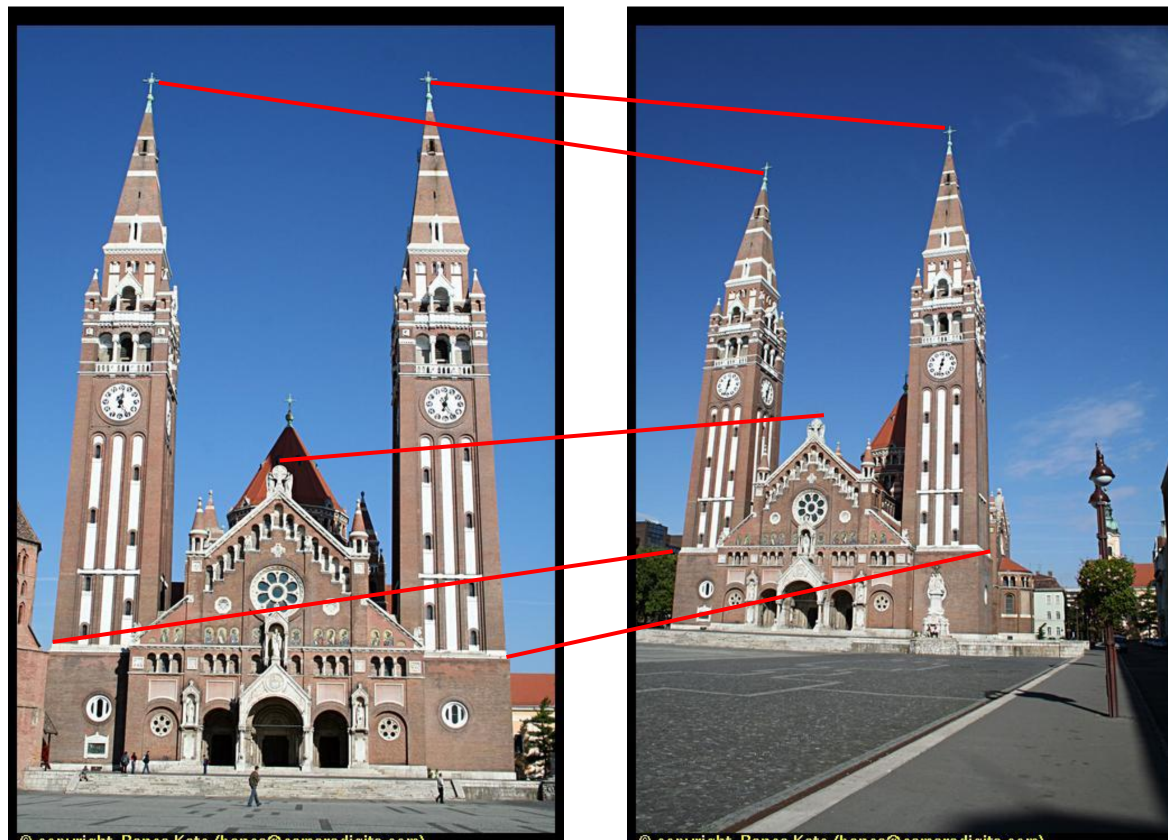
4. Jellemző pontok kinyerése és megfeleltetése

Kató Zoltán

Képfeldolgozás és Számítógépes Grafika tanszék
SZTE
(http://www.inf.u-szeged.hu/~kato/teaching/)

Jellemzők és megfeleltetésük

- **A képfeldolgozás, számítógépes látás számos területén felmerülő probléma egy látványról készült képpár közötti pontmegfeleltetések megkeresése.**
  - **Ehhez megbízható jellemzőket (tipikusan sarokpontokat) nyerünk ki a képekről**
  - **Majd a kinyert pontokat megfeleltetjük egymásnak.**



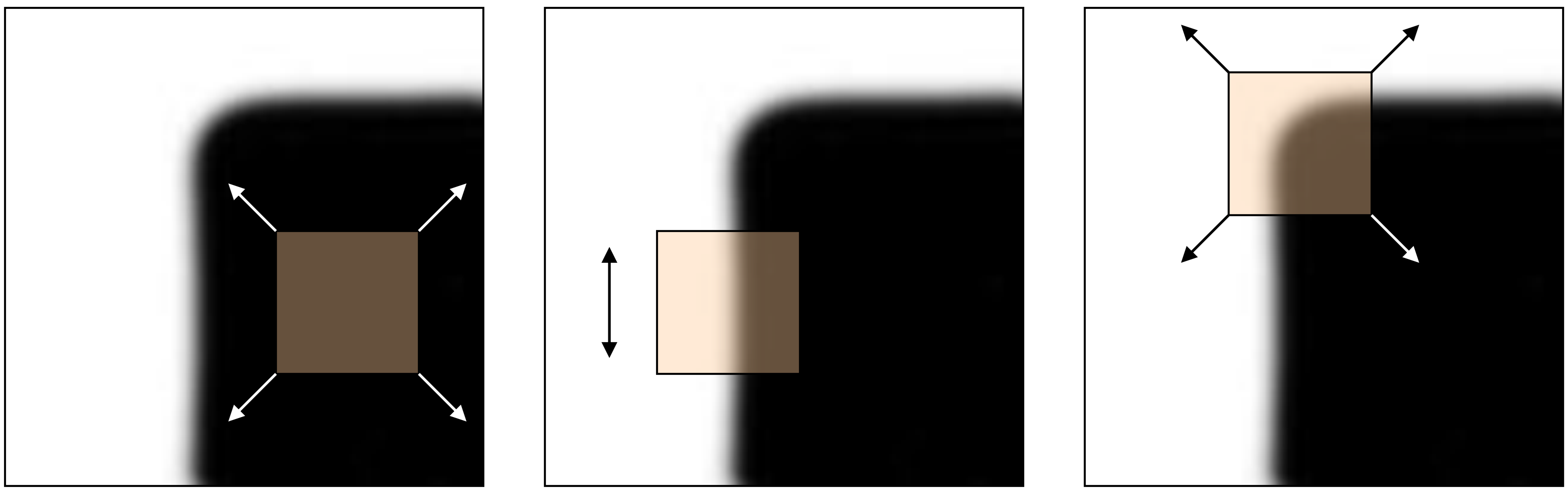
Megválaszolandó kérdések

- Melyek azok a pontok, amelyeket megbízható módon detektálhatunk a képeken?
  - Sarokpontok
- Hogyan tudjuk leírni/jellemezni a kinyert pontokat?
  - Invariáns jellemzők
- Hogyan feleltessünk meg két képről kinyert pontokat?
  - Jellemzők összehasonlítása, robusztusság

Kató Zoltán: Ipari Képfeldolgozás

Sarokpont detektálás

- **Egy csúszóablakot vizsgálva ott lesz sarokpont, ahol az ablakot bármilyen irányba mozgatva nagy képfüggvény változást tapasztalunk**



"sima" régió: egyik irányban sincs változás

"él": az él irányában nincs változás

"sarok": jelentős változás minden irányban

Source: A. Efros

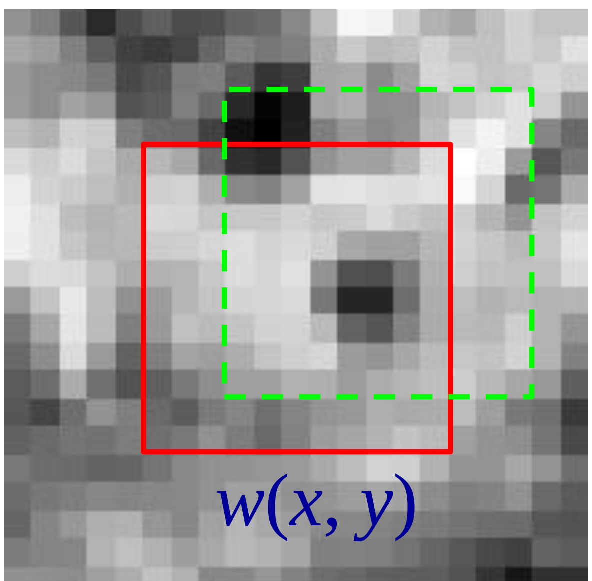
Kató Zoltán: Ipari Képfeldolgozás

Sarokpont detektálás: matematikai modell

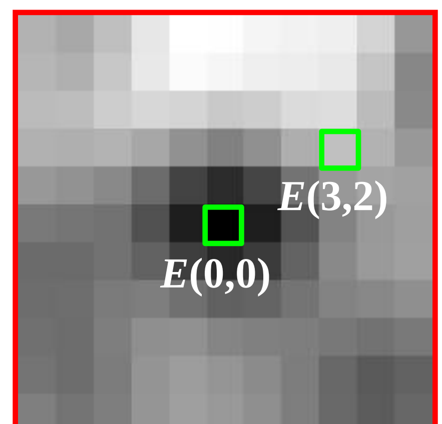
- **A *w*(*x*,*y*) ablak tartalmának változása egy [*u*,*v*] eltolás hatására:**
  - **Lényegében auto-korrelációt számolunk**

$$E(u,v)=\sum_{x,y} w(x,y)\left[I(x+u,y+v)-I(x,y)\right]^2$$

$I(x, y)$

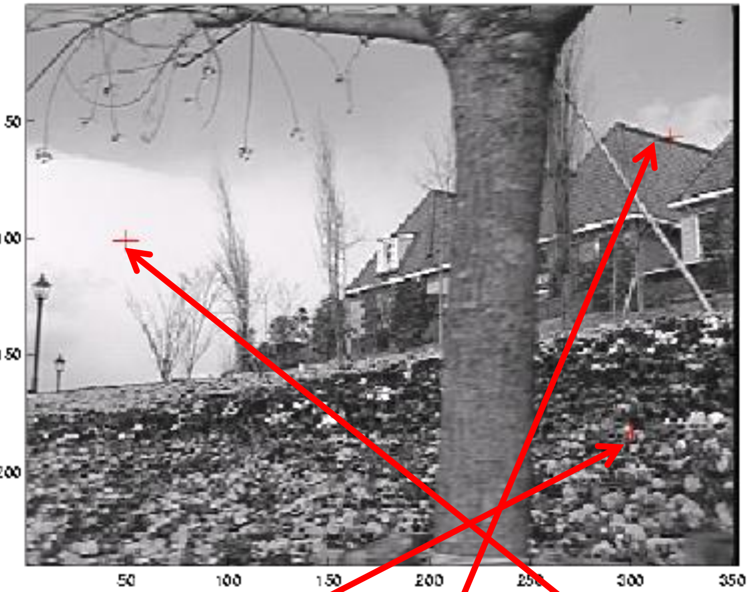


$E(u, v)$

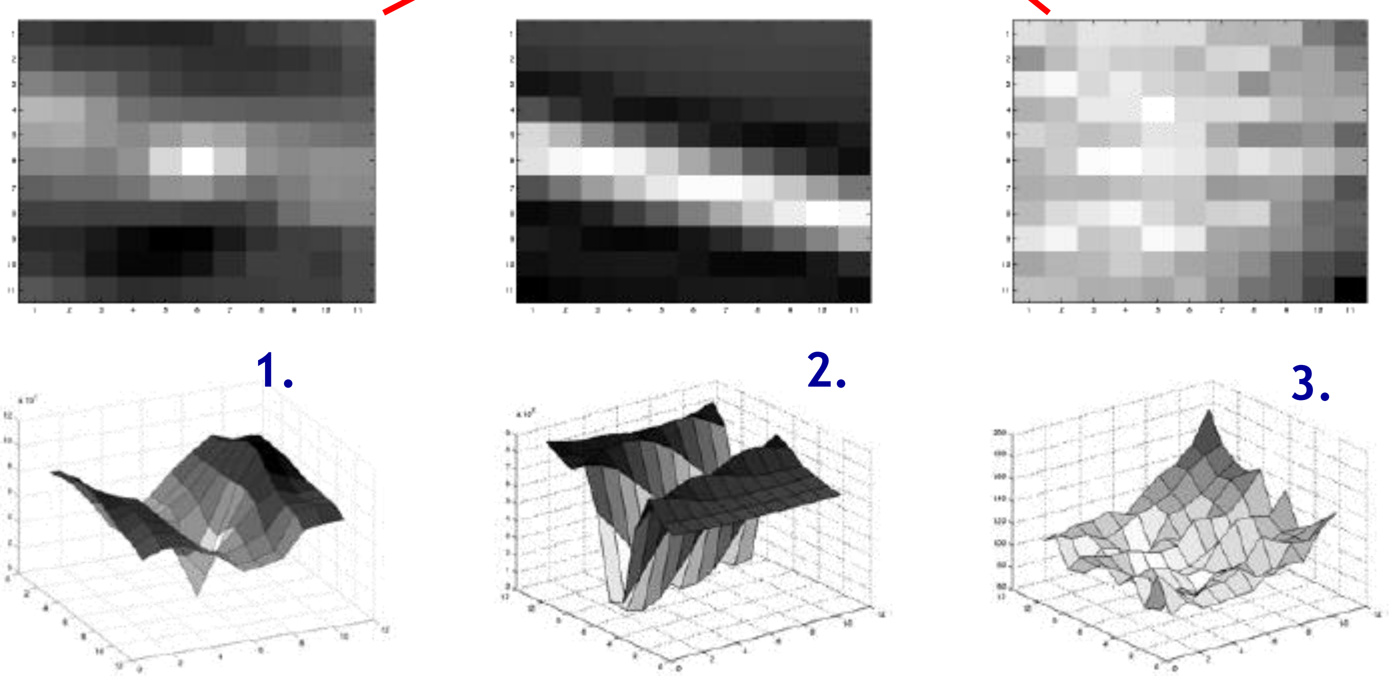


$E(u,v)$ változása

1. **Megfelelően változatos intenzitástartalom ➔jó lokális minimum**
2. **Élek mentén csak az egyik irányban van lokalis minimum**
3. **Homogén területen $E(u,v)$ alig változik**



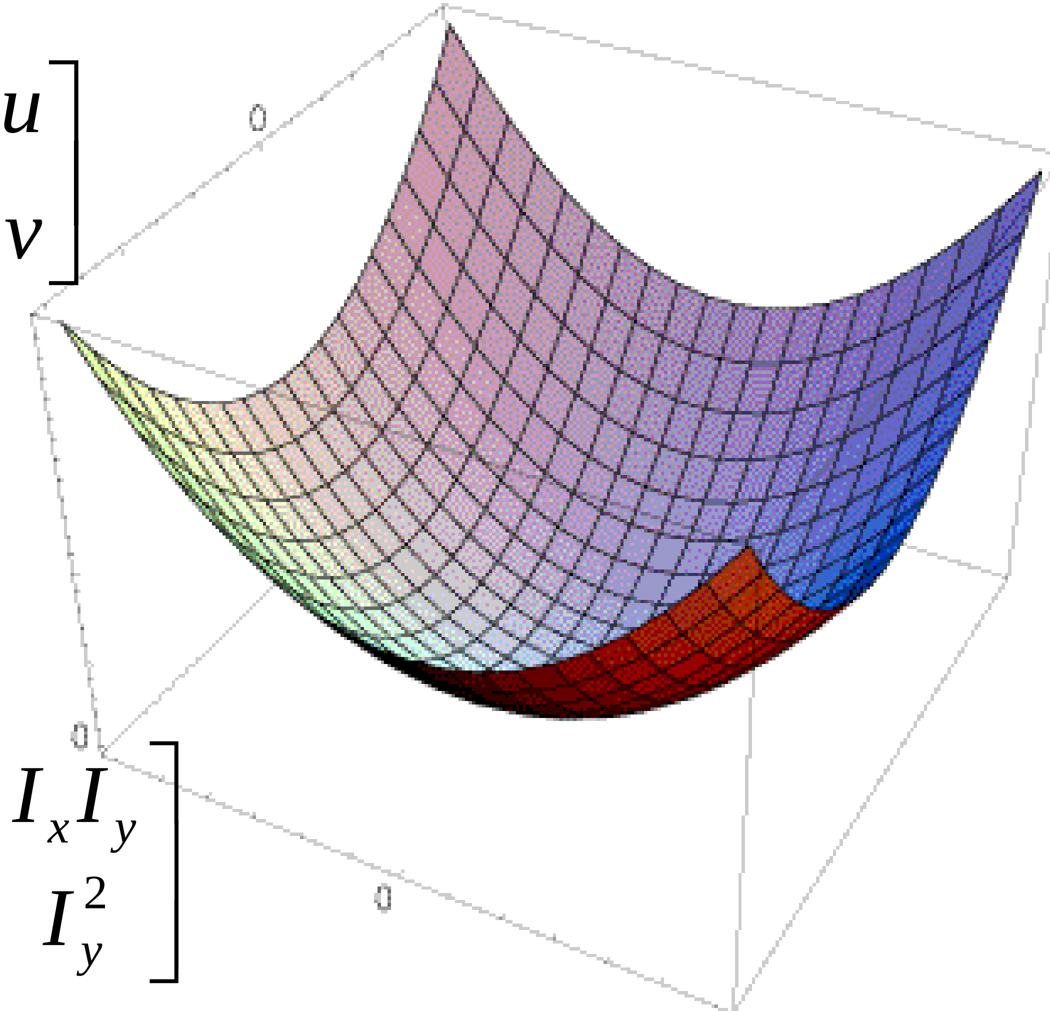
(a)



Az M második momentum mátrix

- **Az *E*(*u*,*v*) felszín lokális approximációja kvadratikus alakban**

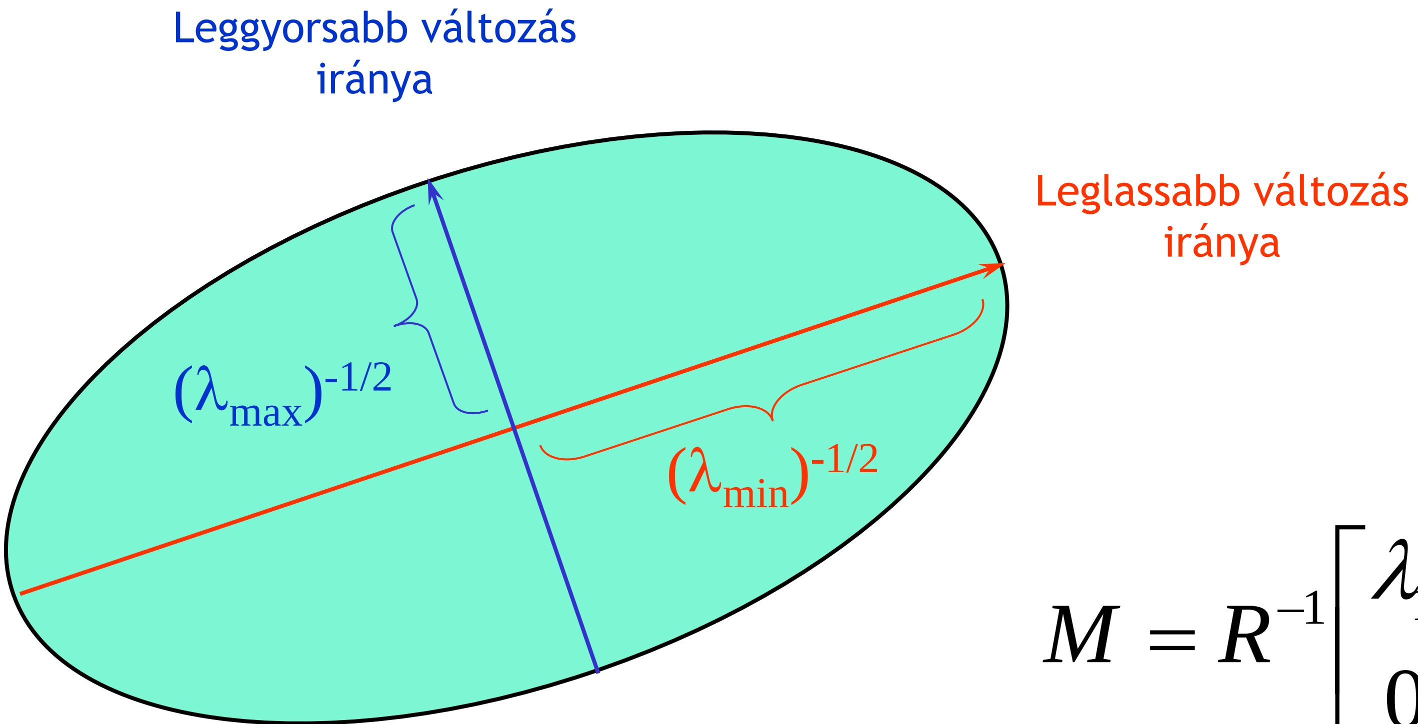
$$E(u,v) \approx [u\ \ v]\ M \begin{bmatrix} u \\ v \end{bmatrix}$$



$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

M mint ellipszis jellemzése

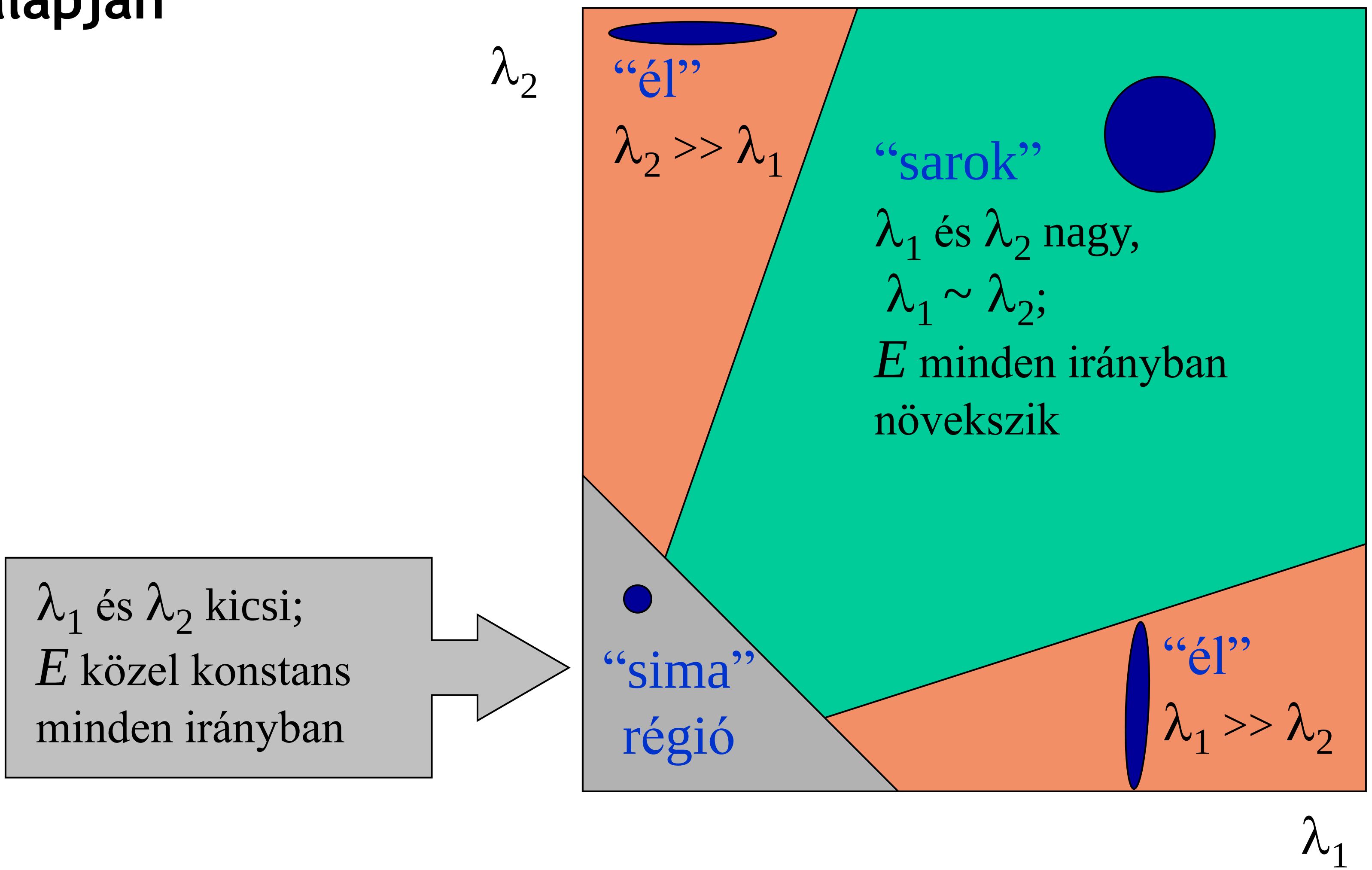
- A tengelyek hosszát a sajátértékek adják
- A tengelyek irányát pedig az *R* forgatási mátrix adja



$$M = R^{-1}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$$

A sajátértékek jelentése

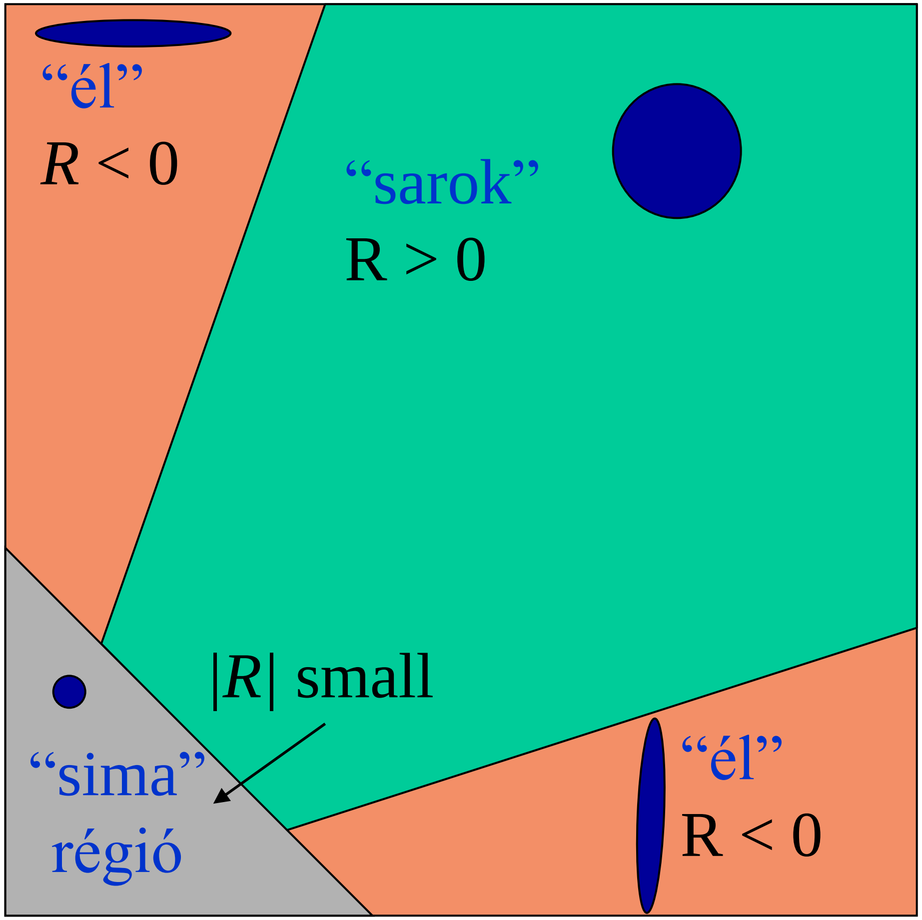
- **A képünk pontjait osztályozhatjuk M sajátértékei alapján**



Sarkosságot jellemző függvény

$$R = \det(M) - \alpha\, \mathrm{trace}(M)^2 = \lambda_1\lambda_2 - \alpha(\lambda_1 + \lambda_2)^2$$

α: konstans (0.04 - 0.06)

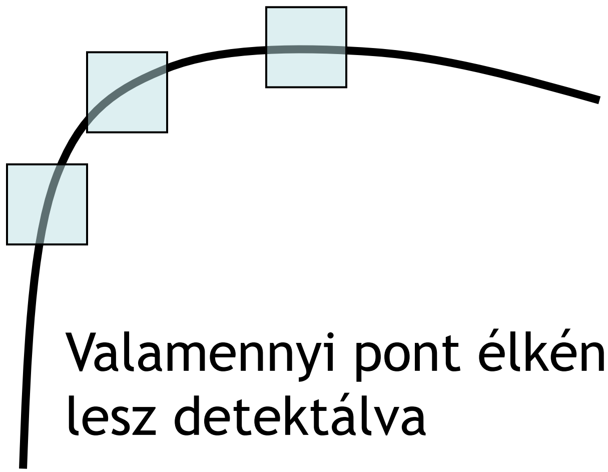


Harris sarokdetektáló algoritmus

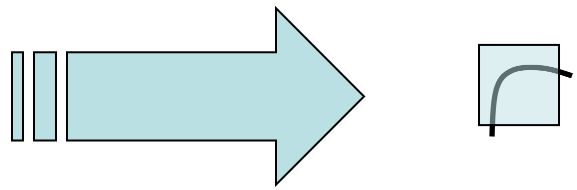
1) **Számítsuk ki az *M* mátrixot minden egyes képpont feletti ablakban és ebből megkapjuk az *R* sarkossági jellemzőt**
2) **Keressük meg azokat a pontokat, amelyekre a sarkossági érték elegendően nagy (*R > küszöb*)**
3) **Tartsuk meg ezekből a lokális maximumokat (vagyis nyomjuk el a nem-maximumokat)**

C.Harris and M.Stephens. "A Combined Corner and Edge Detector." *Proceedings of the 4th Alvey Vision Conference*: pages 147—151, 1988.

Harris: skálázásra nem invariáns



Valamennyi pont élként lesz detektálva

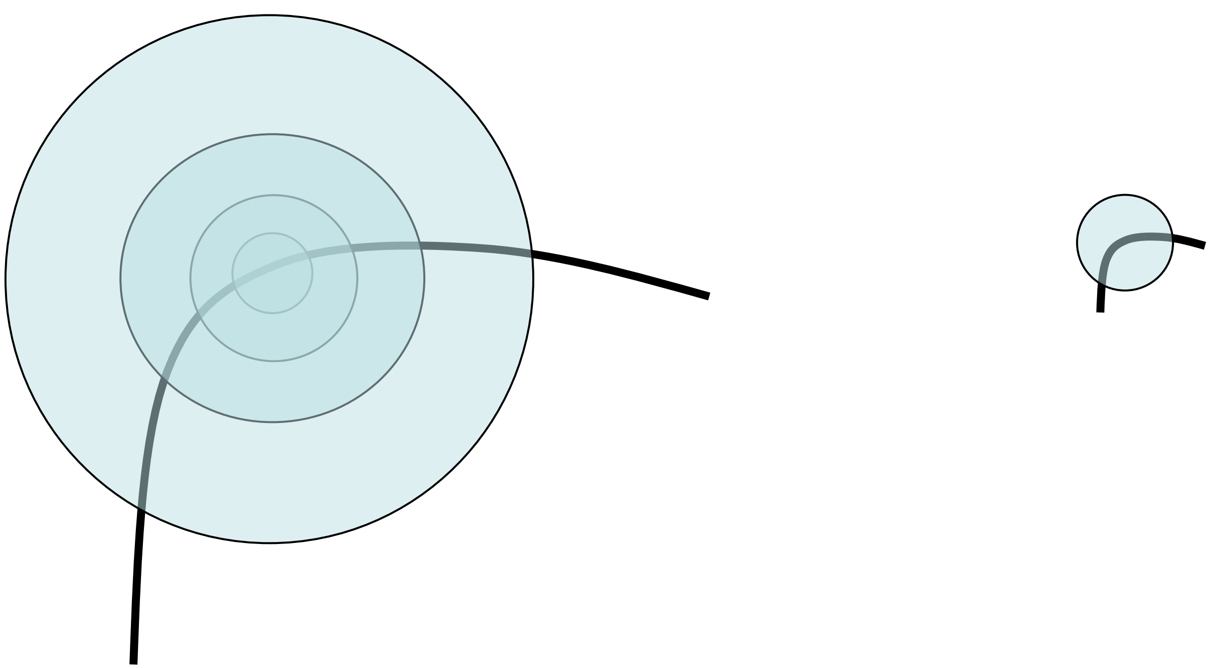


Sarokpont !

- **Hogyan detektálhatunk skála-invariáns, egymásnak megfeleltethető sarokpontokat?**

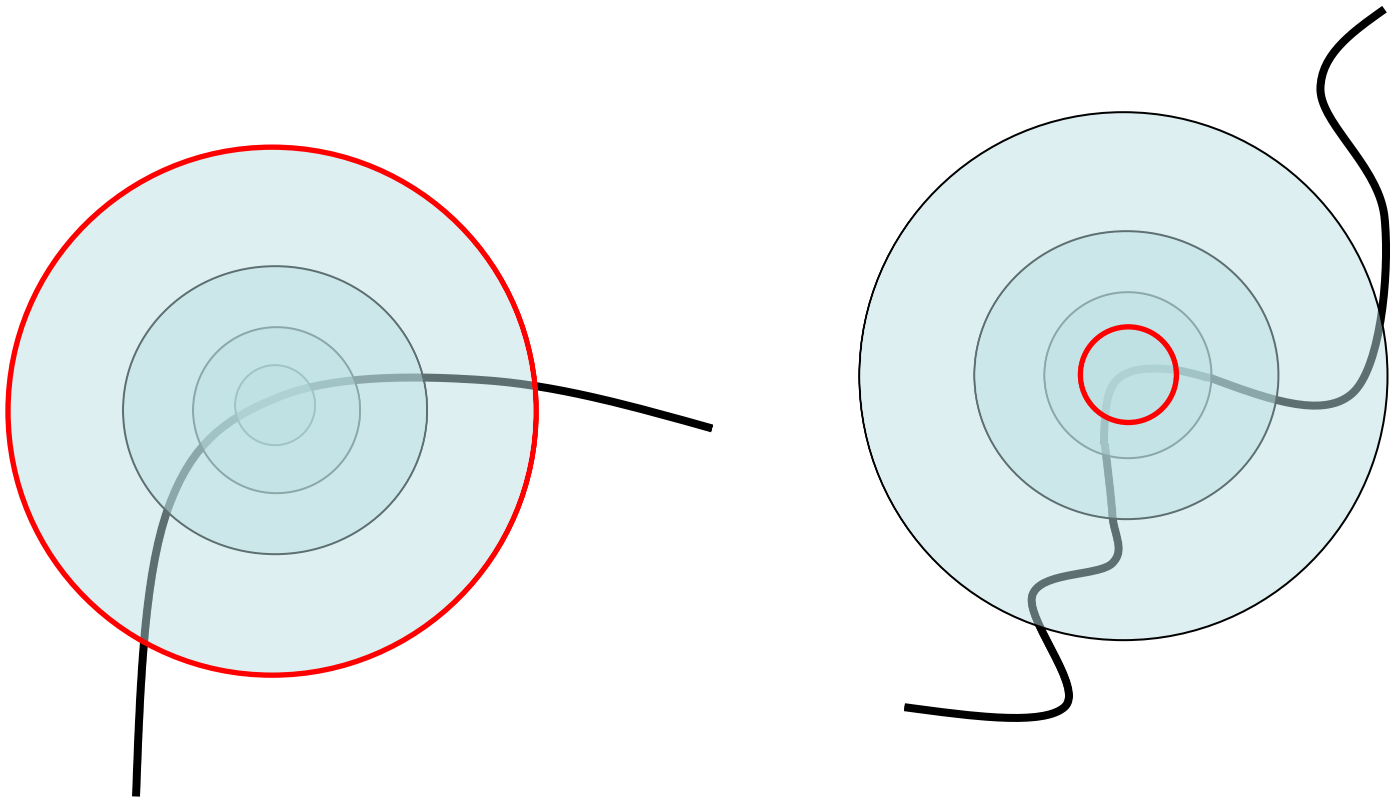
Skála-invaráns detektálás

- Tekintsünk különböző méretű régiókat (pl. köröket) egy pont körül
  - Ezzel ekvivalens, ha megfelelő képpiramis változó felbontású szintjein azonos mérettel keresünk
- Az egymásnak megfelelő méretek hasonlóan néznek ki mindkét képen



Skála-invaráns detektálás

- Hogyan válasszuk ki ***egymástól függetlenül*** a "megfelelő" köröket?

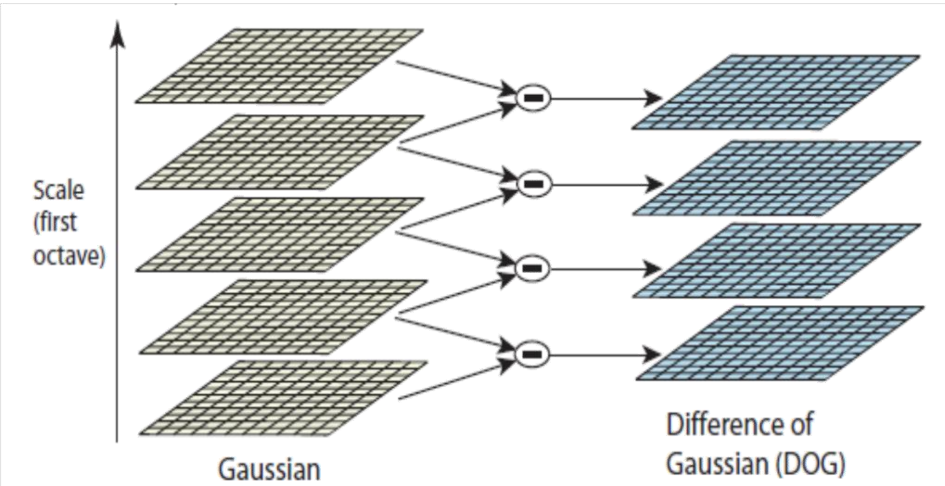


Skála invariáns sarok-detektálás

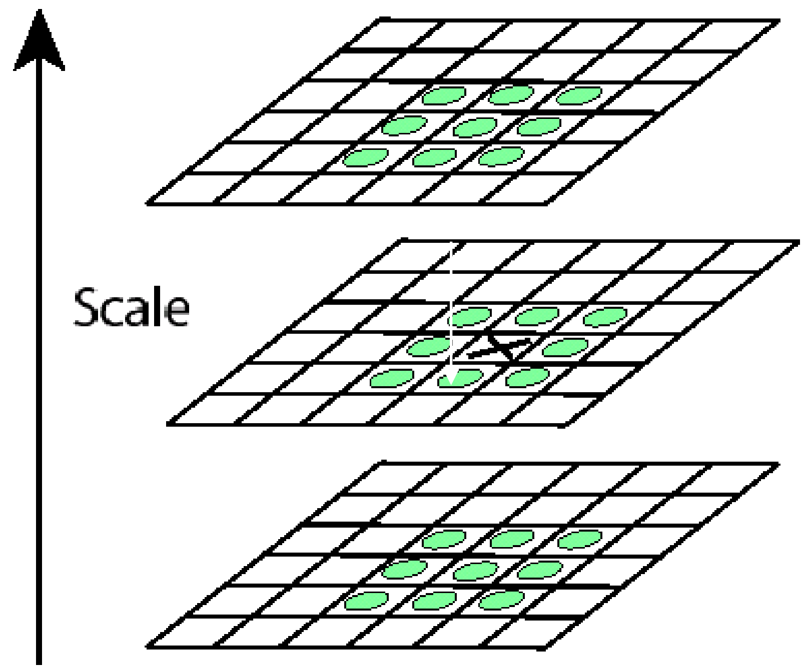
- **Input:** **két kép ugyanazon látványról nagy felbontásbeli különbséggel**
- **Cél:** **detektáljuk *ugynazon* jellemző pontokat mindkét képen *függetlenül* a képmérettől**
- **Megoldás:** **keressük a *maximumát* egy megfelelően konstruált függvénynek a *skála-* és *képtérben***

Jellemző pontok kinyerése DoG segítségével

- **Detektáljuk a difference-of-Gaussian (DoG) szélsőhelyeit skála-térben**
  - **(max, min) egy 3*3*3 szomszédságban**

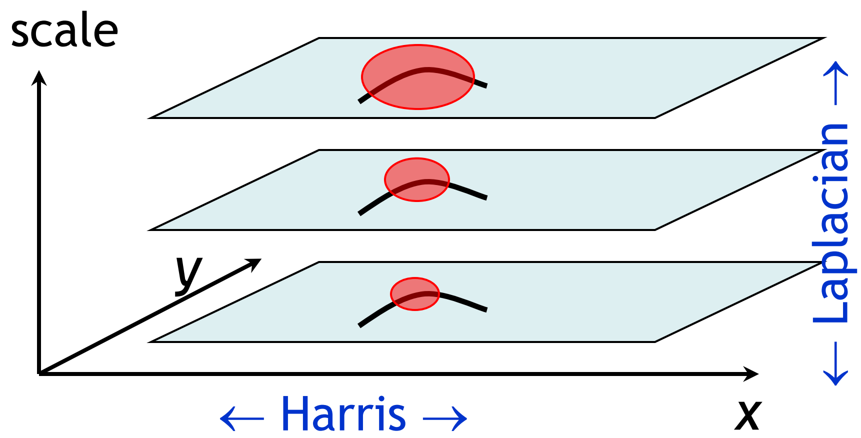


- **Küszöböljük az értékeket**
- **Elimináljuk az él-válaszokat**
- **Az így kapott pontok listája: (x,y,σ)**

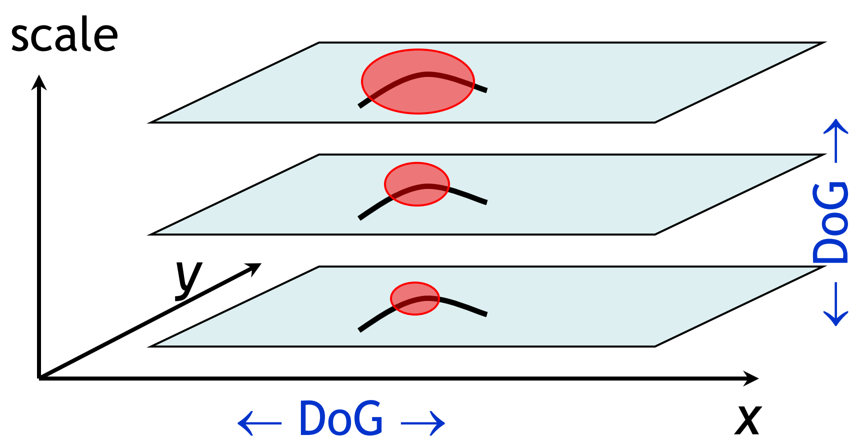


Skála invariáns sarok-detektálás

- **Harris-Laplacian**[1]
  Keressük meg a lokális maximumait:
  - **Harris sarokdetektor a képen**
  - **Laplacian a skálatérben**



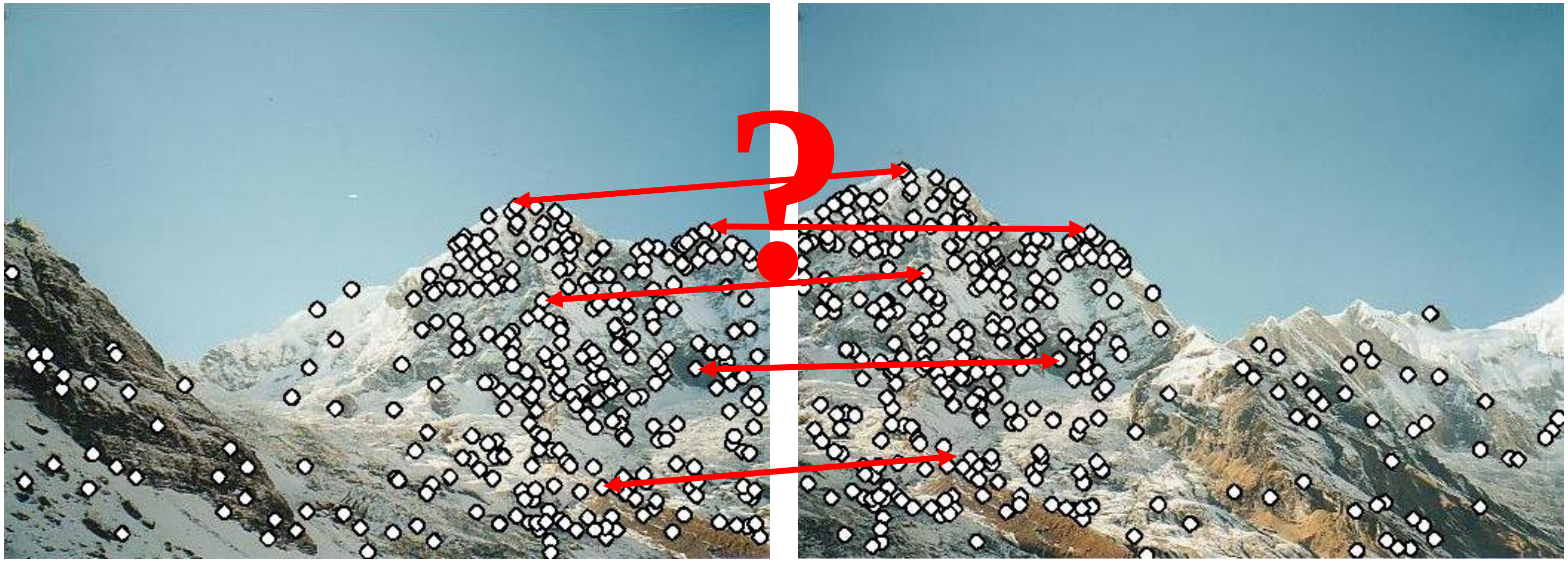
- **SIFT (Lowe)**[2]
  Keressük meg a lokális maximumait:
  - **Difference of Gaussians (DoG) a skála- és képtérben**



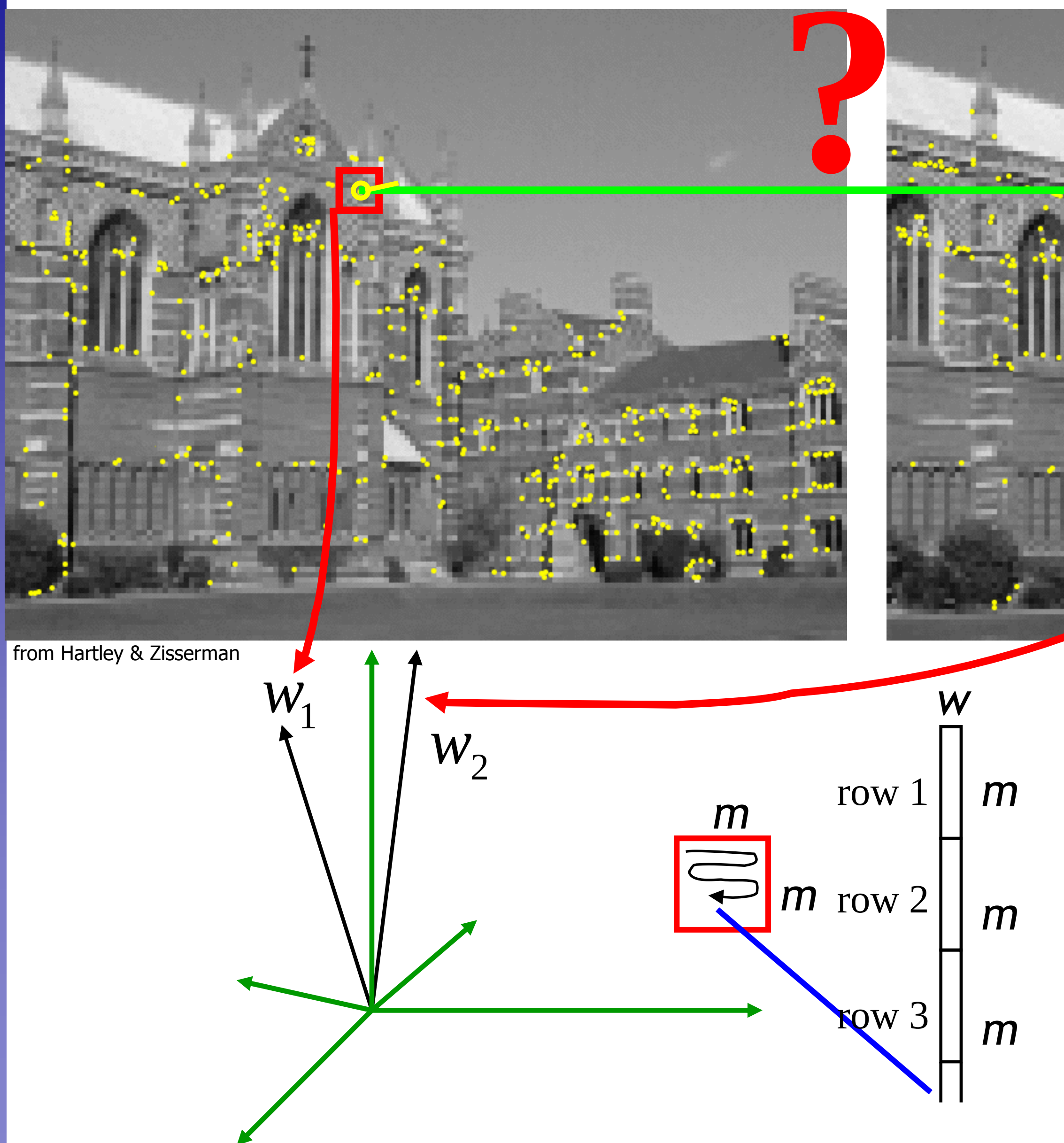
[1] K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004

Lokális leírók

- A detektált pontokat hogyan tudnánk leírni úgy, hogy az
  - invariáns és
  - egyedi legyen
- A kinyert pontok önmagukban nem jellemezhetők jól
  - egyetlen intenzitás-érték nem elég stabil és egyedi
  - A pontok környezete már elég egyedi lehet, de az invariancia biztosítása nem triviális
- Tekintsük a pontot tartalmazó ablak tartalmát



Képfoltok mint vektorok



Mindkét *mxm* ablak tartalma ábrázolható egy *m*² dimenziós vektorként.

Az ablakok tartalmát normalizálva a vektorok egységnyiek lesznek, miáltal könnyen összehasonlíthatóak a két vektor különbségeként.

Az ablakok tartalmának normalizálása

- Még azonos kamerák esetén is lehet eltérés a rögzített intenzitás értékekben (pl. eltérő gain/érzékenység).
- Ennek kiegyenlítésére célszerű a képeket normalizálni
  - Ezáltal az ablakokból képzett vektorok egységnyiek lesznek, így különböző képek között is összehasonlíthatóak

$$\bar{I} = \frac{1}{|W_m(x,y)|} \sum_{(u,v)\in W_m(x,y)} I(u,v)$$

Átlag pixelérték

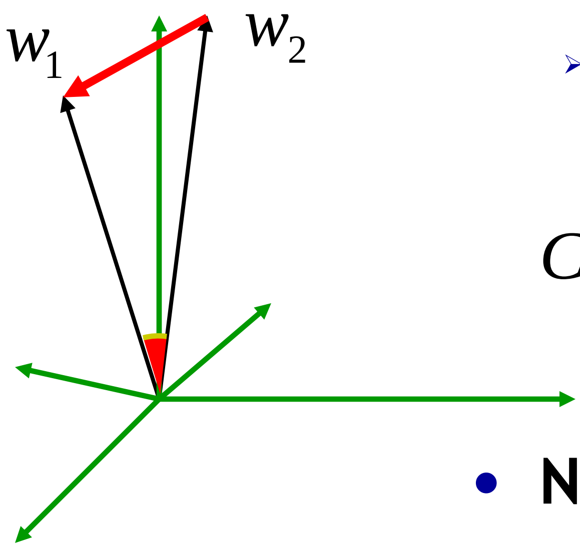
$$\|I\|_{W_m(x,y)} = \sqrt{\sum_{(u,v)\in W_m(x,y)} [I(u,v)]^2}$$

Ablak magnitúdó

$$\hat{I}(x,y) = \frac{I(x,y) - \bar{I}}{\|I - \bar{I}\|_{W_m(x,y)}}$$

Normalizált pixelérték

Hasonlósági mértékek



- **SSD: (Normalized) Sum of Squared Differences**
  - **Minél kisebb annál jobban illeszkedik a két vektor**

$$C_{\mathrm{SSD}} = \sum_{(u,v)\in W_m} [\hat{I}_1(u,v) - \hat{I}_2(u,v)]^2 = \left\| w_1 - w_2 \right\|^2$$

- **NC: Normalized Correlation**
  - **Minél nagyobb, annál jobban illeszkedik a két vektor**

$$C_{\mathrm{NC}} = \sum_{(u,v)\in W_m} \hat{I}_1(u,v)\hat{I}_2(u,v) = w_1 \cdot w_2 = \cos\theta$$

Képfolt alapú leíró invarianciája

- Fotometriai invariancia a normalizálással elég jól teljesül
- Azonos méretű ablakok hasonlíthatók össze ➔ skálafüggő
- Azonos állású (vagyis rasztersorokra illeszkedő) ablakok hasonlíthatók össze ➔ orientáció függő

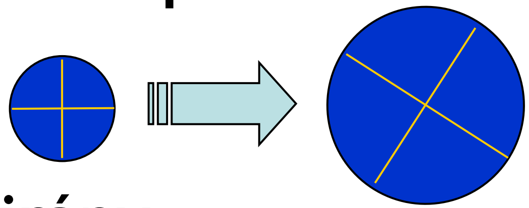
- Összességében tehát
  - nincs geometriai invariancia,
  - van részleges fotometriai invariancia

SIFT: SCALE INVARIANT FEATURE TRANSFORM

SIFT – Scale Invariant Feature Transform

- Skála- és irányfüggetlen fotometriailag invariáns pont-leírókat állít elő az alábbi főbb lépésekben:

1. **skála** meghatározása (ez már megtörténik a pontok detektálása során)
   - DoG szélsőhelyek térben és skálában

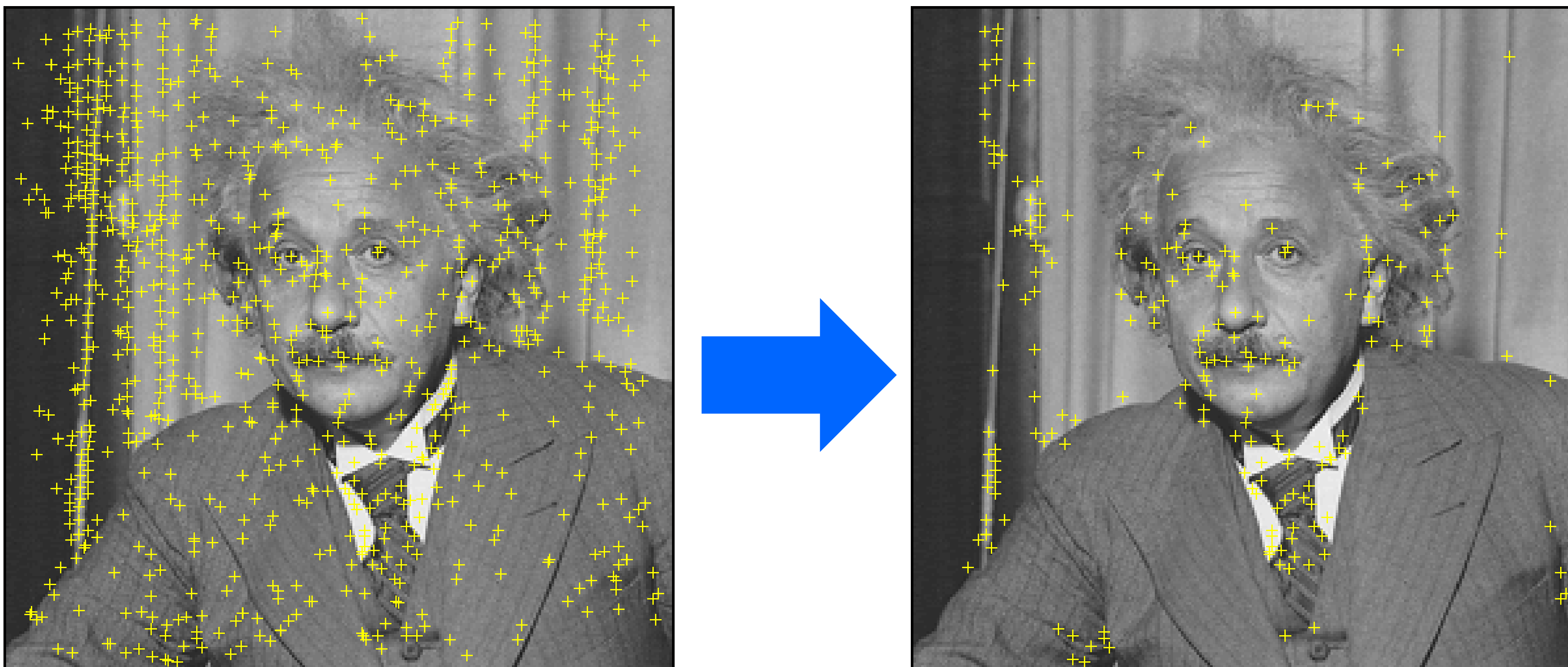


2. **lokális orientáció** a domináns gradiens irány
3. Az így kapott skála és orientáció minden egyes kinyert pontban egyértelműen meghatároz egy lokális koordinátarendszert
   - Minden további számítás ebben a koordinátarendszerben történik, így a kapott leírók skála- és irányfüggetlenek lesznek
4. Számítsunk **gradiens irány-hisztogramokat** több kisebb ablakban, amiből leíró vektort képezünk.

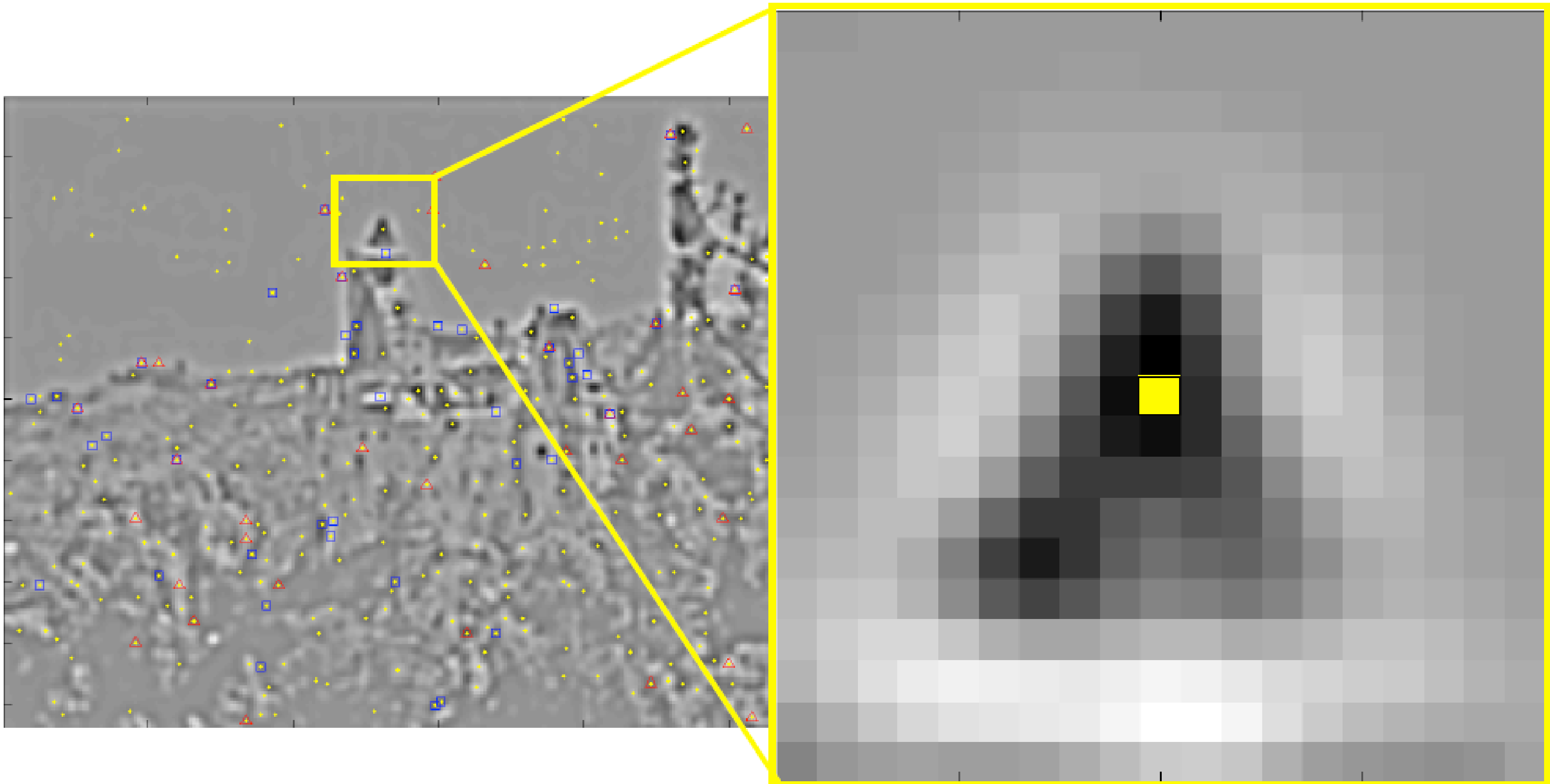
D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. *International Journal of Computer Vision*, 2004

Jellemző pontok kinyerése

- DoG skálatérben válasszunk ki minden pontot, ami a 3x3x3 környezetben szélsőhely (max, vagy min)
- Töröljük az instabil pontokat
  - Alacsony kontraszt
  - Nem elég magas sarkossági jellemző
- Megkapjuk a képi pozíciót (x,y) és a hozzá tartozó skálát (σ)

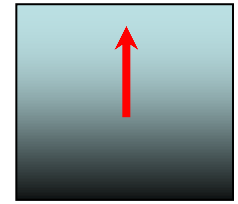


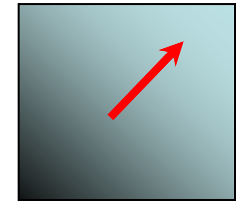
Kinyert jellemző pont



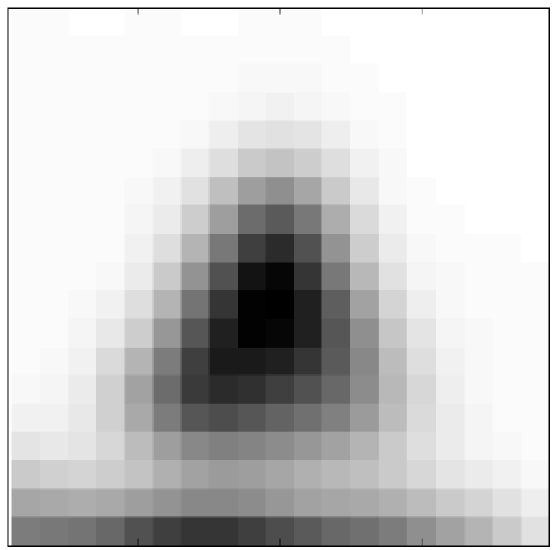
•Keypoint location = extrema location
•Keypoint scale is scale of the DOG image

Orientáció



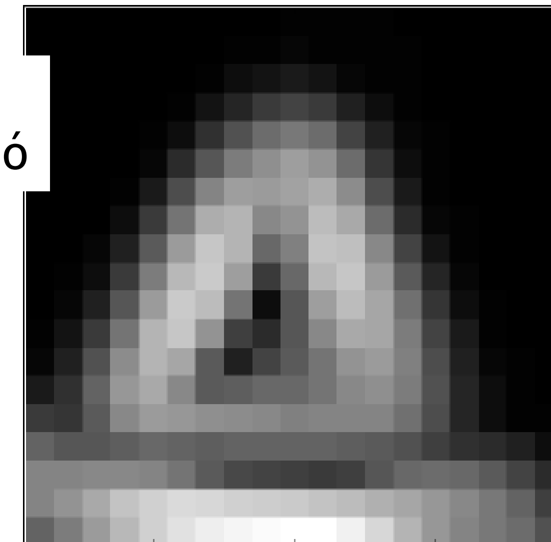


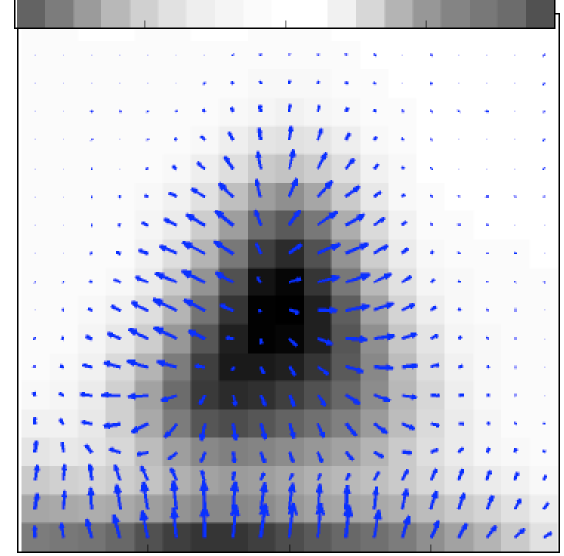
- **A kinyert pontokhoz rendeljünk egy konzisztens irányt**
  - **Minden ponthoz a hozzá tartozó skálán számoljunk gradienst**



A képpiramis megfelelő skálájáról vett (Gauss simított) kép

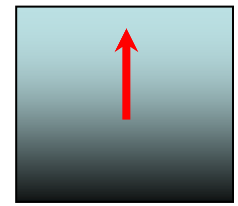
gradiens magnitúdó

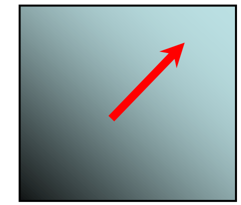




Gradiens irány

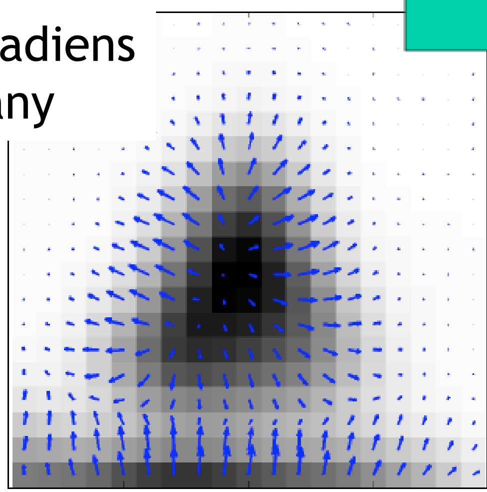
Orientáció



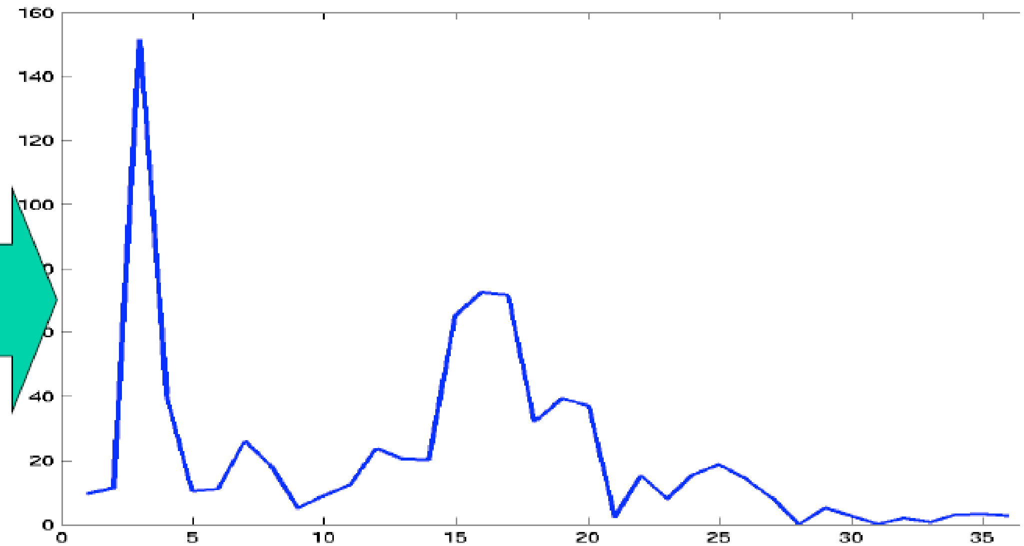


Súlyozott gradiens magnitúdó

Gradiens irány

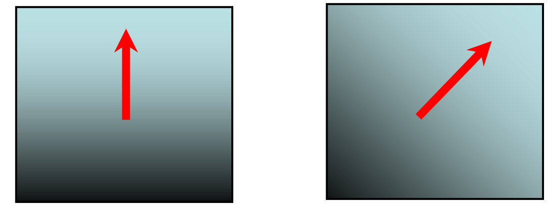


Súlyozott irány-hisztogram: Minden csoport azon gradiens-magnitúdók súlyozott összegét tartalmazza, amelyek a csoportba eső irányúak



Irány csoportok: 10 fokonként tekintsük a gradiens vektorokat. Ez összesen 36 csoportot jelent.

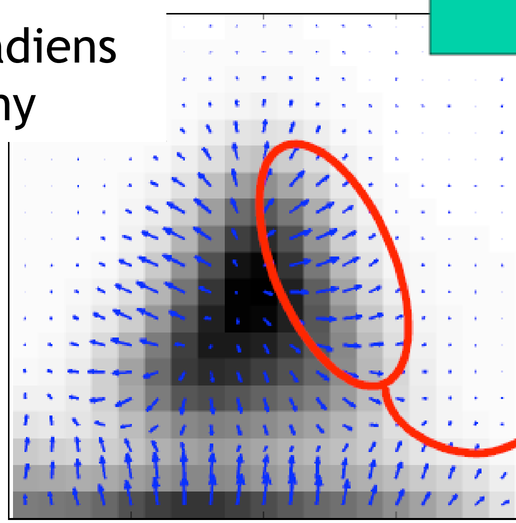
Orientáció

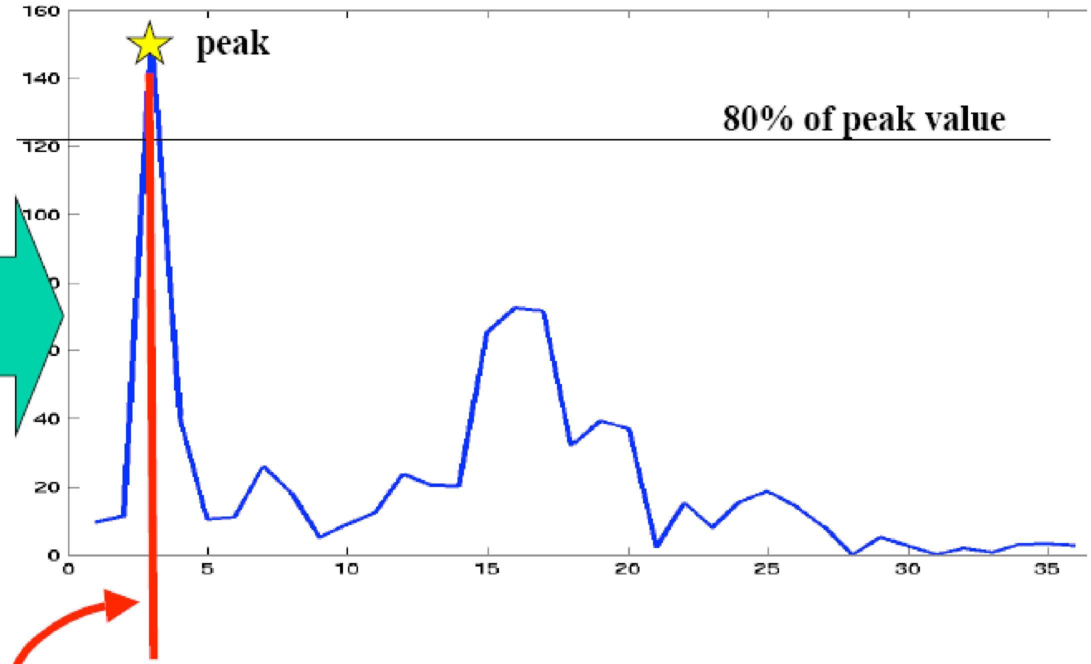


Súlyozott gradiens magnitúdó

Irány kiválasztás: Minden olyan irányt kiválasztunk, amelyik a csúcsérték 80%-n felül van és minden ilyen irányhoz létrehozunk egy pontot (tipikusan a pontok 15%-a több irányú)

Gradiens irány

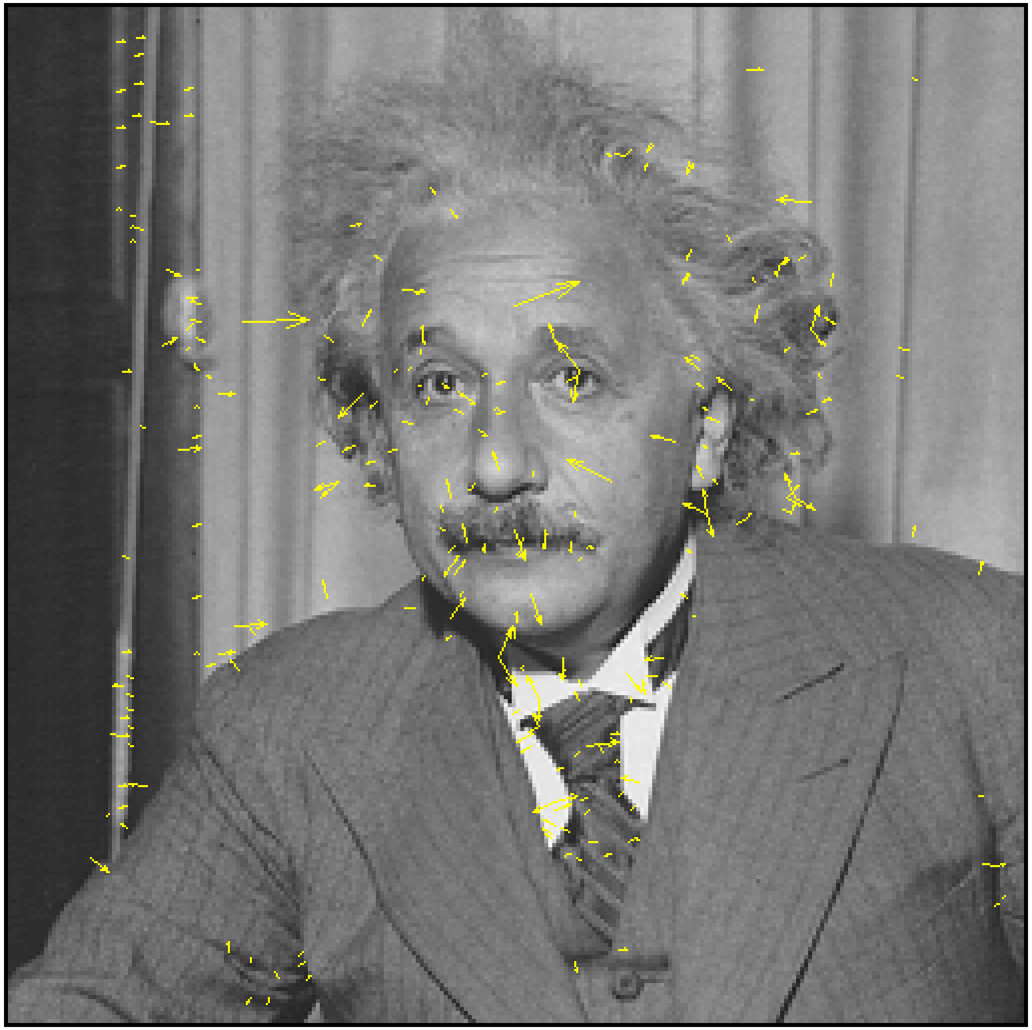




Jelen példában a pont iránya 25 fok lesz, mivel csúcs a 20-30 fokos csoportban van, és nincs másik domináns irány.

SIFT leíró

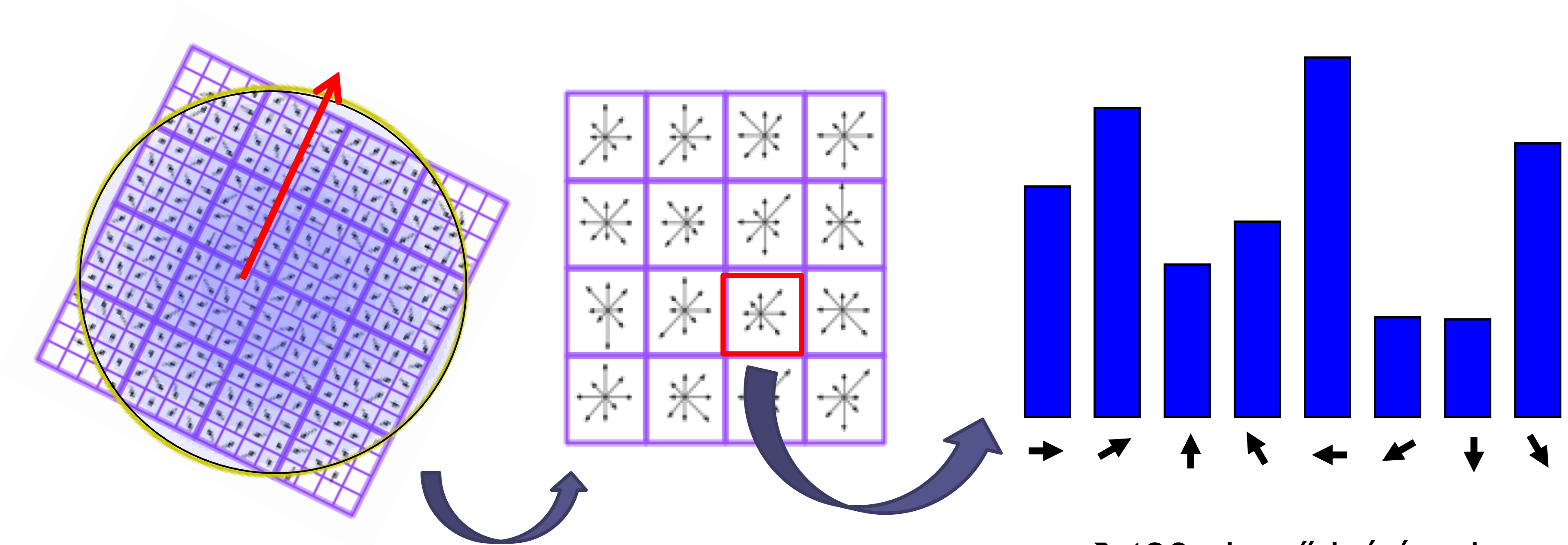
- Eddig minden jellemző ponthoz tartozik:
  - pozíció
  - skála
  - magnitúdó
  - irány



- Mivel írhatnánk le a pont körüli régiót?

SIFT leíró

- Tekintsünk egy 16x16 ablakot a gradiens-mezőn, amit tovább osztunk 4x4 blokkokra.
- Számítsunk a 4x4-es mintákon 8 irányban hisztogramot
- Alkalmazzunk Gauss súlyozást a középpont körül, aminek szórása=0.5xσ (a jellemző pont skálája)
- 4x4x8 = 128 dimenziós vektor leíró



→128 elemű leíró vektor

SIFT leíró – fotometriai invariancia

- Lineáris (globális) megvilágítás változás:
  - Az intenzitásértékek nagysága (gain) önmagában nem befolyásolja a gradienst
  - A vektorok egységnyivé normalizálása feloldja a kontrasztbeli különbségeket
- Nem lineáris (lokális) megvilágítás változás:
  - Telítettség inkább a magnitúdót mint az irányt befolyásolja.
  - Küszöböljük a gradiens magnitúdót 0.2-vel ([0,1]-re normált értékeket feltételezve), majd normalizáljunk újra.
    - empirikusan meghatározott érték 3D objektum többféle megvilágítása alapján
- Az legnépszerűbb, rendkívül robusztus pontleíró [Mikolajczyk & Schmid 2005]

Krystian Mikolajczyk and Cordelia Schmid: GLOH “A performance evaluation of local descriptors”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp1615—1630, 2005

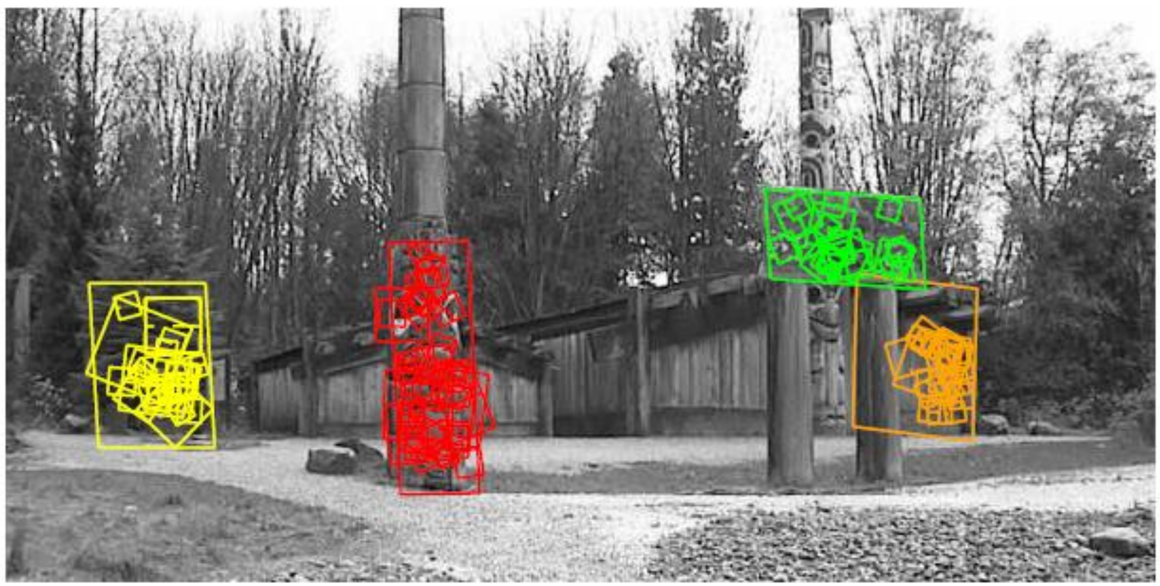
Kató Zoltán: Ipari Képfeldolgozás

SIFT leírók megfeleltetése

- L2 távolság alapján a legközelebbi szomszéddal párosítjuk
- Hogyan szűrjük ki a rossz megfeleltetéseket (pl. ha nincs megfelelő párja egy pontnak)?
  - Küszöböljük az L2 távolságot ➔ rossz teljesítmény
  - Küszöböljük inkább az arányt ➔ jó teljesítmény

$$\frac{\text{legjobb megfeleltetés}}{\text{2. legjobb megfeleltetés}}$$

Képrészletek felismerése



Regisztráció panoráma képhez

[Brown & Lowe 2003]

Felhasznált anyagok

- Trevor Darrell: C280, Computer Vision
  http://www.cs.cmu.edu/afs/cs/academic/class/15385-s06/lectures/ppts/
- David Lowe
- James Hays: CS 143 Computer Vision, Brown University
  - http://www.cs.brown.edu/courses/cs143/
- További források az egyes diákon megjelölve