

Invariáns képjellemzők detektálása és követése

Orientáció és skálázás invariáns jellemzők detektálása és követése

Darya Frolova, Denis Simakov, David Lowe diáit
is felhasználva

Detektálás és követés - alkalmazások

- Képilelesztés (panoráma)
- 3D rekonstrukció
- Mozgáskövetés
- Alakfelismerés
- Indexálás képi adatbázisokban
- Robot navigáció
- ...

Milyen megfeleltetéseket látunk a példa képsorozaton?

Képzeletben klikkeljünk a képeken összetartozó pontpárookra!
És mindezt algoritmikusan hogyan is lehetne megfogalmazni?



Alkalmazási példák képekkel
illusztrálva:

Képilllesztés



by [Diva Sian](#)



by [swashford](#)

Képilllesztés – ez nehezebb eset

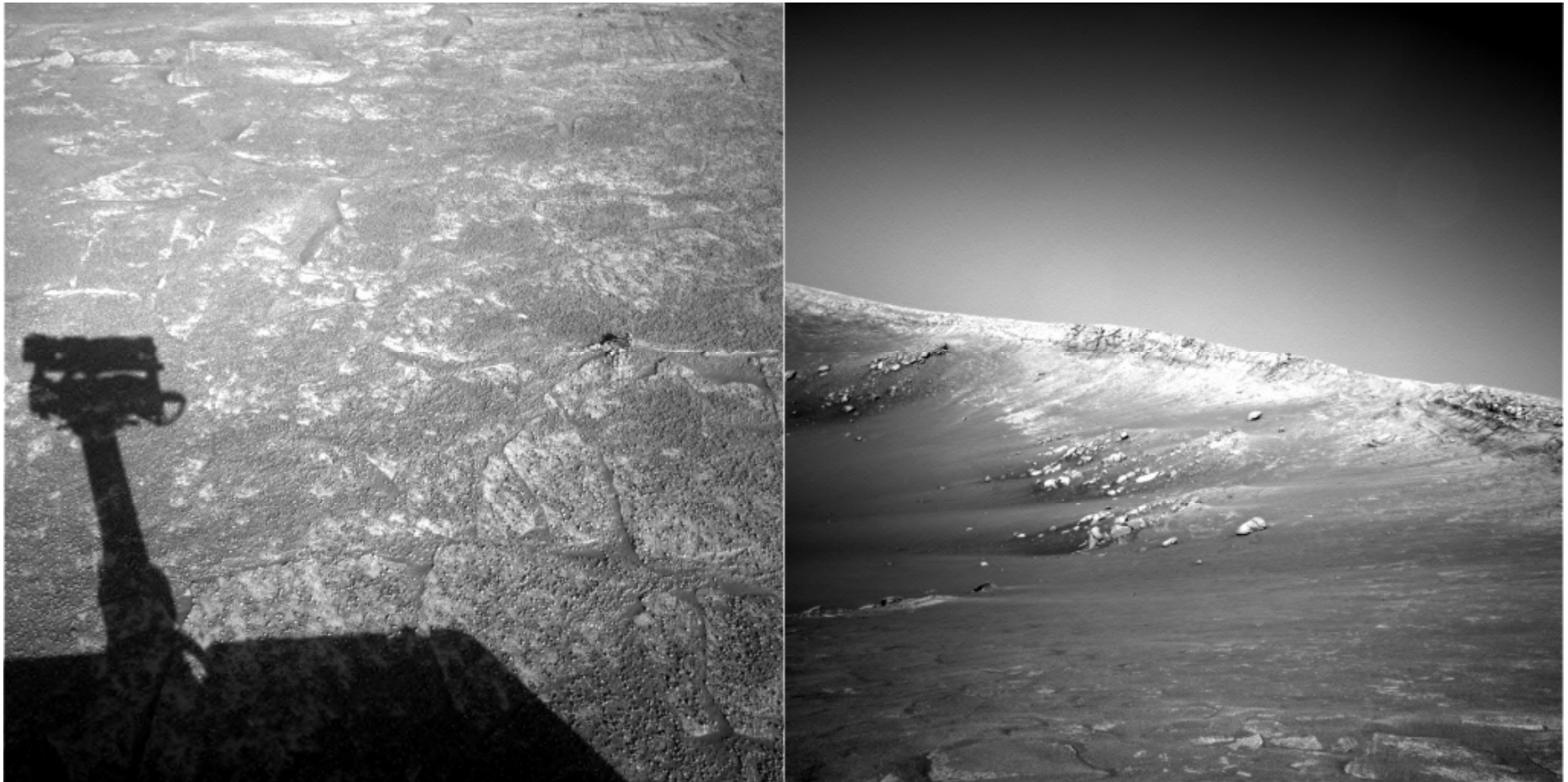


by [Diva Sian](#)



by [scgbt](#)

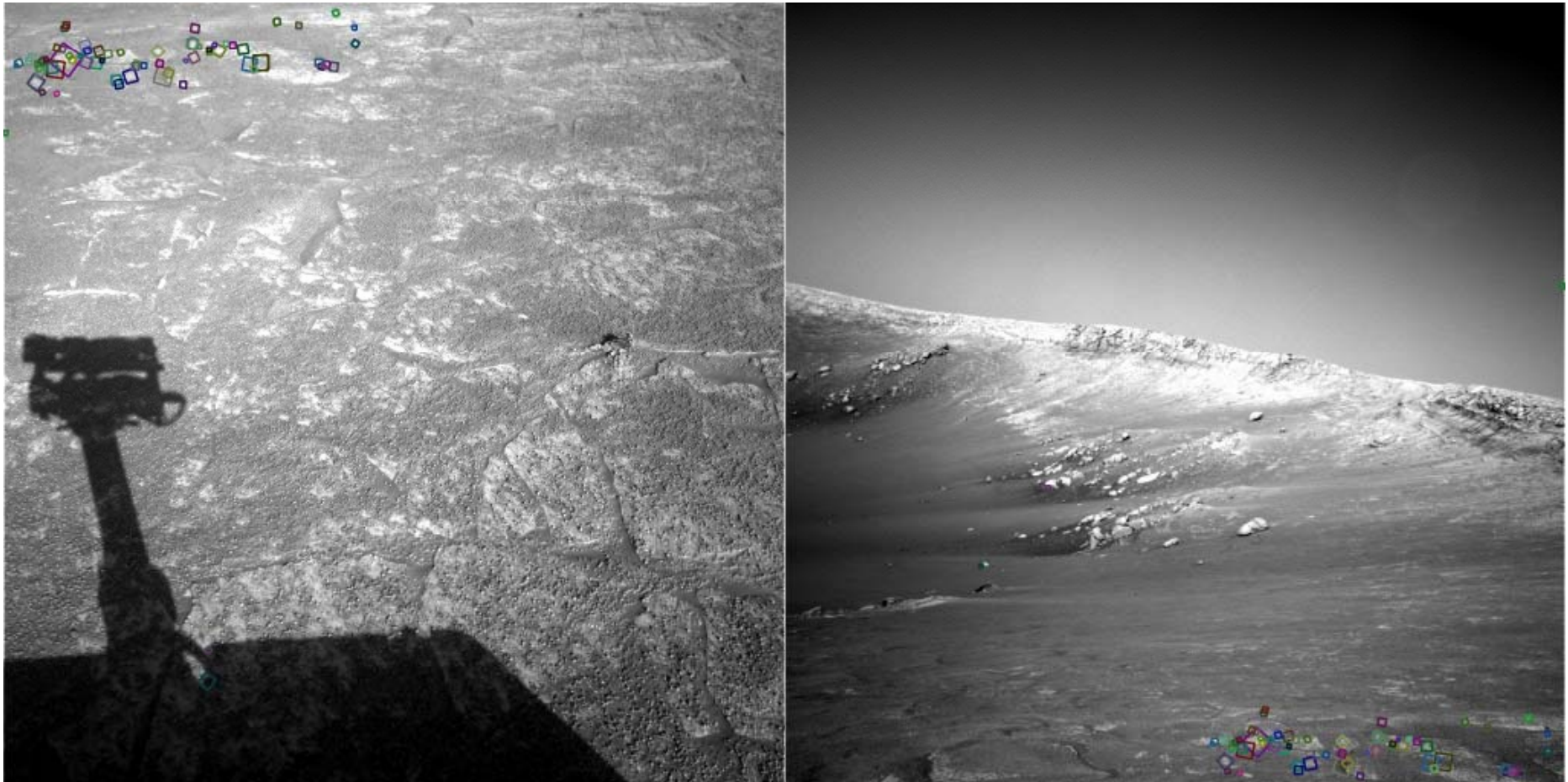
Képilllesztés – ez még nehezebb eset



NASA Mars Rover images

Hogyan illesszünk?

(ld. kis színes négyzeteket)



NASA Mars Rover images
with SIFT feature matches
Figure by Noah Snavely

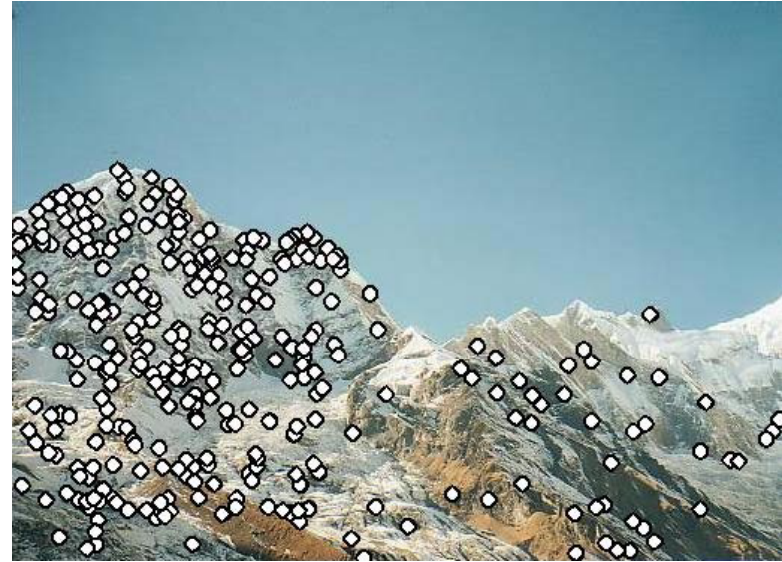
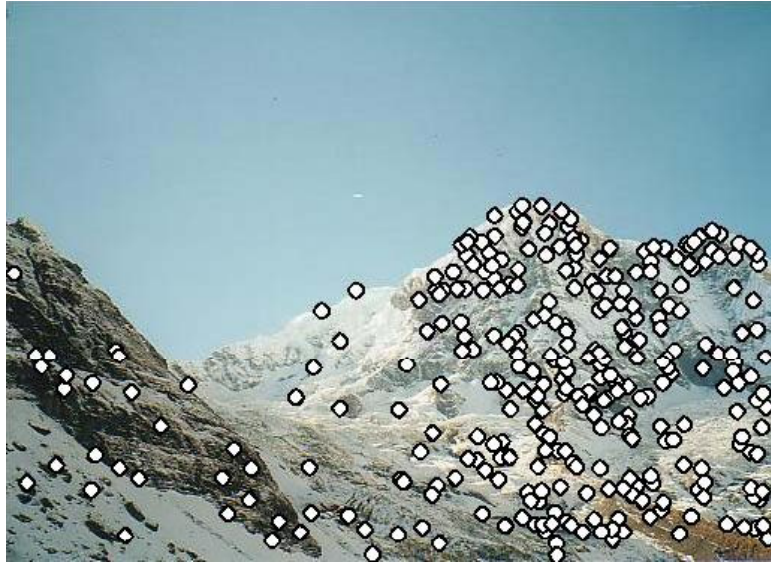
Panorámakép készítés lépései

- Képsorozat illesztés szükséges



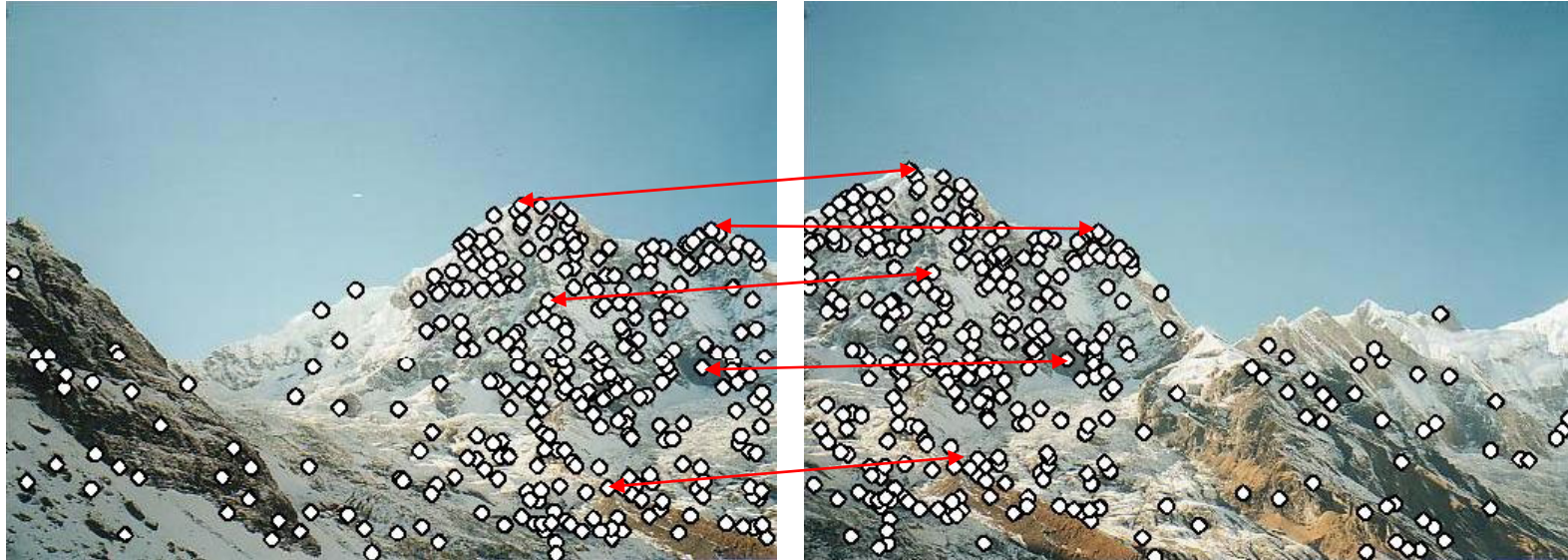
Panorámakép készítés lépései

1.Képjellemző pontok detektálása mindkét képen



Panorámakép készítés lépései

1. Képjellemző pontok detektálása mindkét képen
2. Összetartozó pontpárok megkeresése



Panorámakép készítés lépései

1. Képjellemző pontok detektálása mindkét képen
2. Összetartozó pontpárok megkeresése
3. A megtalált összetartozó pontok segítségével az illesztés (homography)



Hogyan?

(algoritmusok)

Mit akarunk általában?

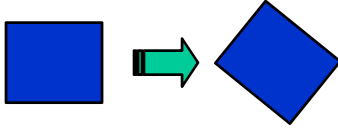
Ugyanazon „érdekes pontok”
detektálása képenként, függetlenül a
képről-képre elszenvedett jelentős
képtorzulásoktól

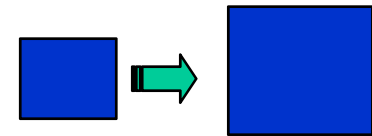
Invarianciát akarunk!

- A „jó” jellemző ellenáll a lehető legtöbb „torzításnak” a képek sorozatán keresztül

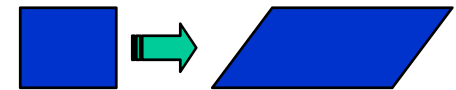
Képtorzulások transzformációs modellje (ismétlés)

- Geometriai torzítás:

- elforgatás 
- hasonlóság (elforgatás + skálázás)



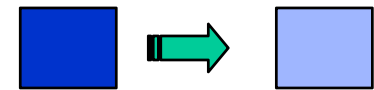
- Affin tr. (irányfüggő skálázás)



- ...

- Fotometriai torzítás:

- Affin intenzitás változás ($I \rightarrow a I + b$)



Tipikus zajhatások

- Megvilágítás



Tipikus zajhatások

- Megvilágítás +
- Skálázás



Tipikus zajhatások

- Megvilágítás +
- Skálázás +
- Elforgatás



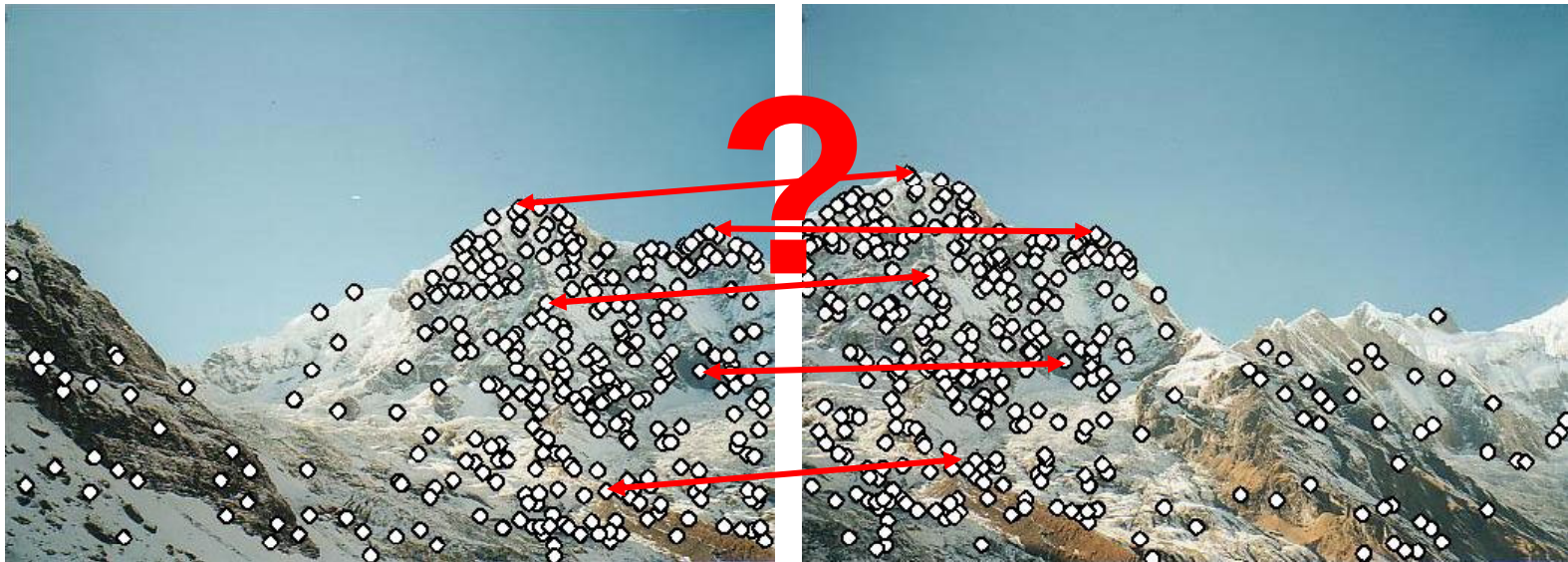
Tipikus zajhatások

- Megvilágítás +
- Skálázás +
- Elforgatás +
- Affin transzformáció +
-



(Jellemző) pont leírás

Ezeket a jellemzőket tehát hogyan illesszük?

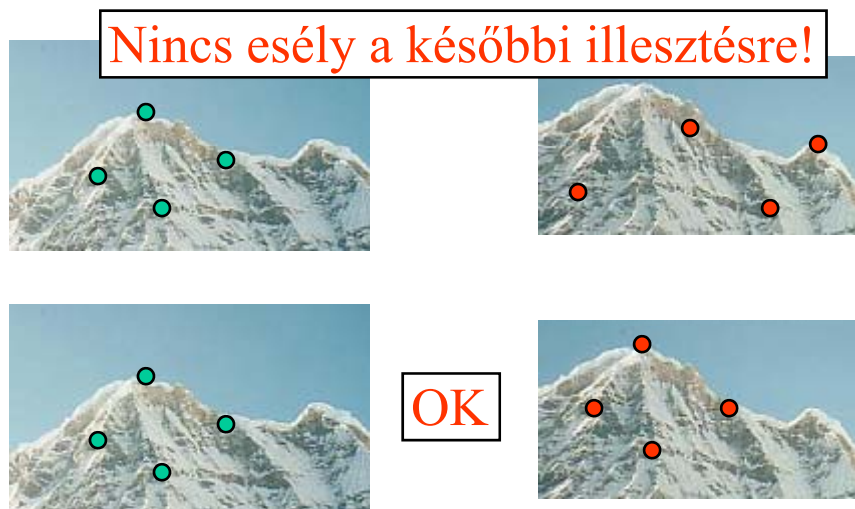


Pont leírás legyen:

1. Invariáns a zajokra
2. Megkülönböztető (egyedi)

Képjellemzők

- 1. probléma:
 - Ugyanazon pontokat függetlenül megtalálni mindkét képen



Képről-képre reprodukálható
hatású **pontdetektor** kellene!

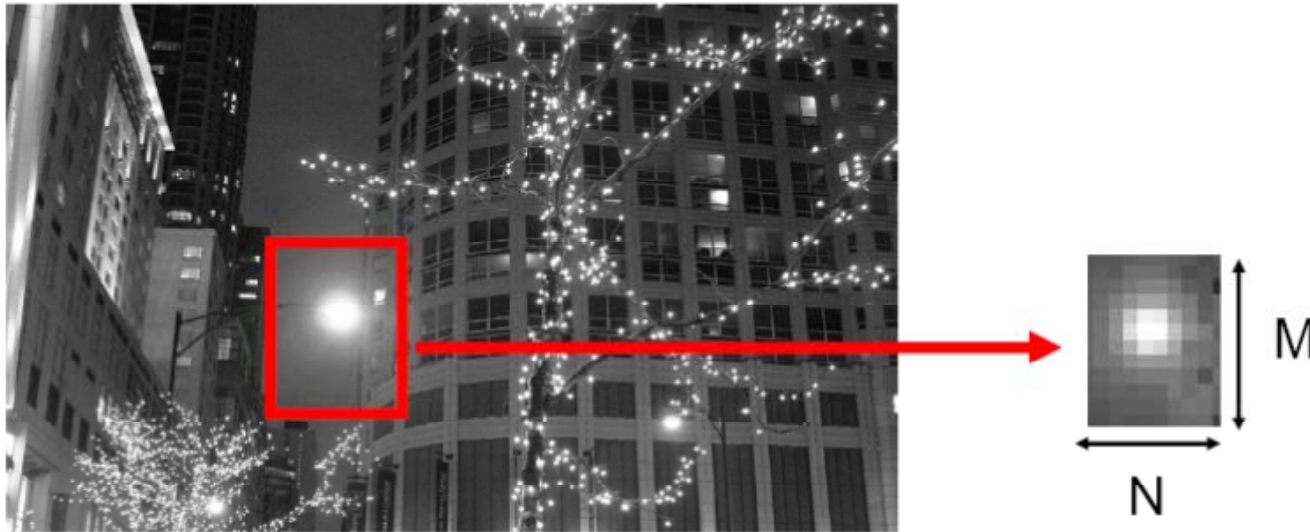
Képjellemzők

- 2. probléma:
 - Minden detektált ponthoz megtalálni az illeszkedőt a következő képen



Képről-képre megbízható és megkülönböztető
hatású **pontleíró** is kellene!

Legegyszerűbb illesztés (eddig) 1: Korrelációval

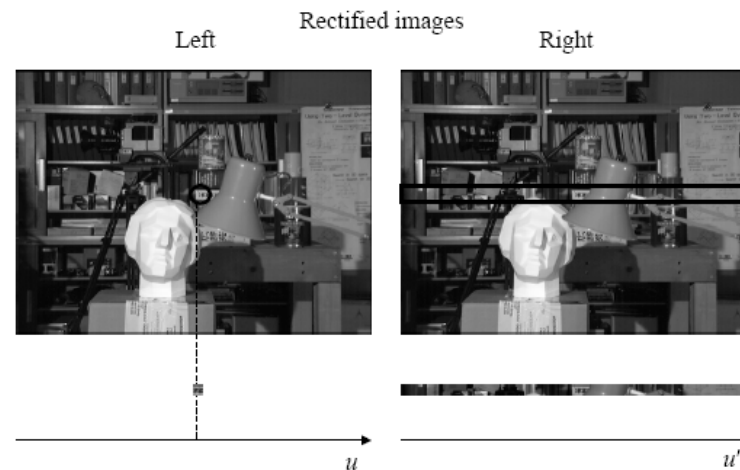


Már többször hivatkoztam erre a megoldásra

Legegyszerűbb illesztés (eddig) 1: Korrelációval

Kielégítően működik ha:

- leginkább csak elmozdulás van
- pl. sztereo, motion sztereo kis mozgásokra



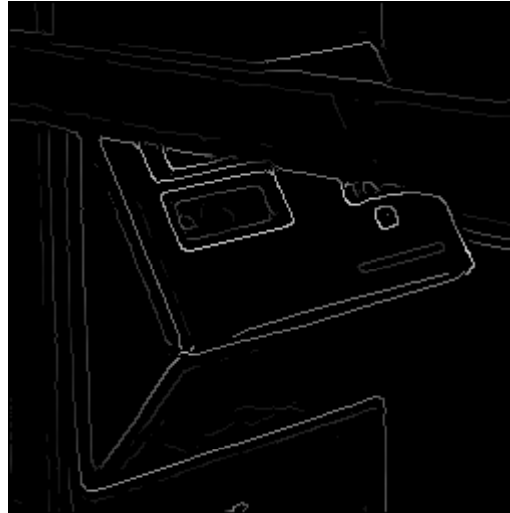
Viszont:

- egyéb torzításokra nem túl invariáns
- Kicsi az egyes kulcspontokat megkülönböztető képessége (egyedisége)

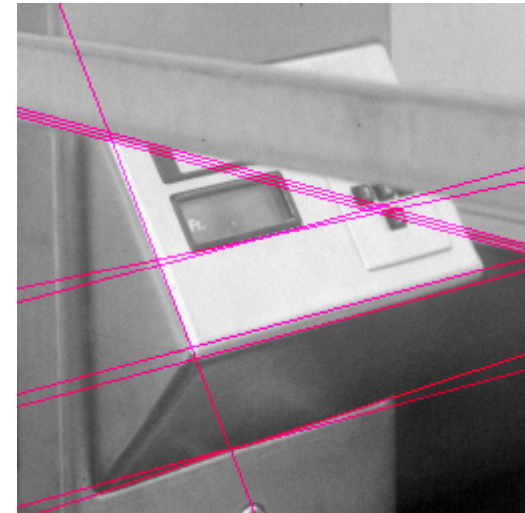
Legegyszerűbb illesztés (eddig) 2: Hough Transzformáció



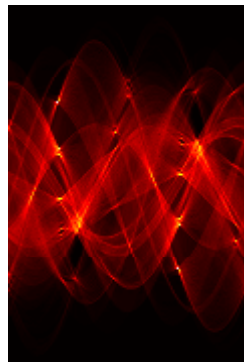
Eredeti
kép



Canny éldetektálás



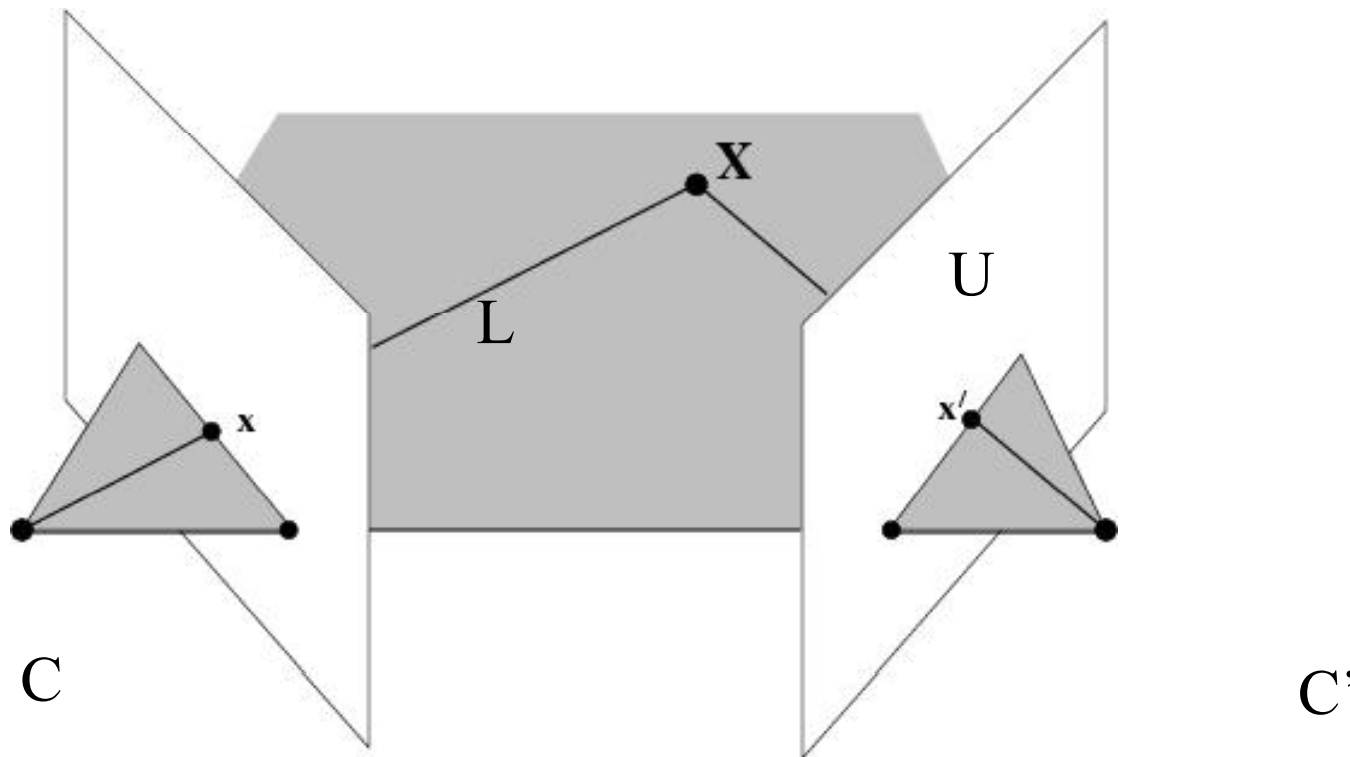
Hough egyenes
keresés



Hough
egyenes-paraméter tér

Már láttuk ezt is...

Legegyszerűbb illesztés (eddig) 3: epipoláris geometria

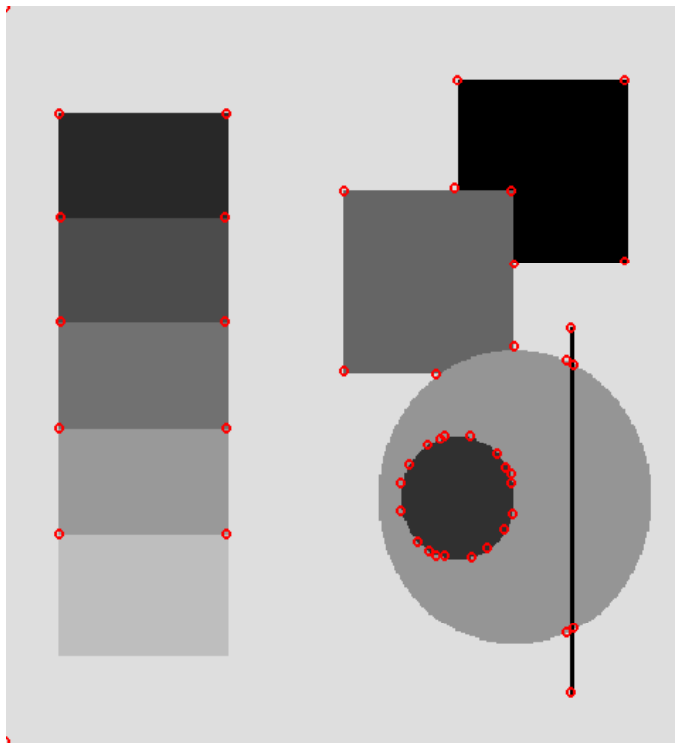


Többnézetű geometriánál majd részletesen,
Most csak az alapelv...

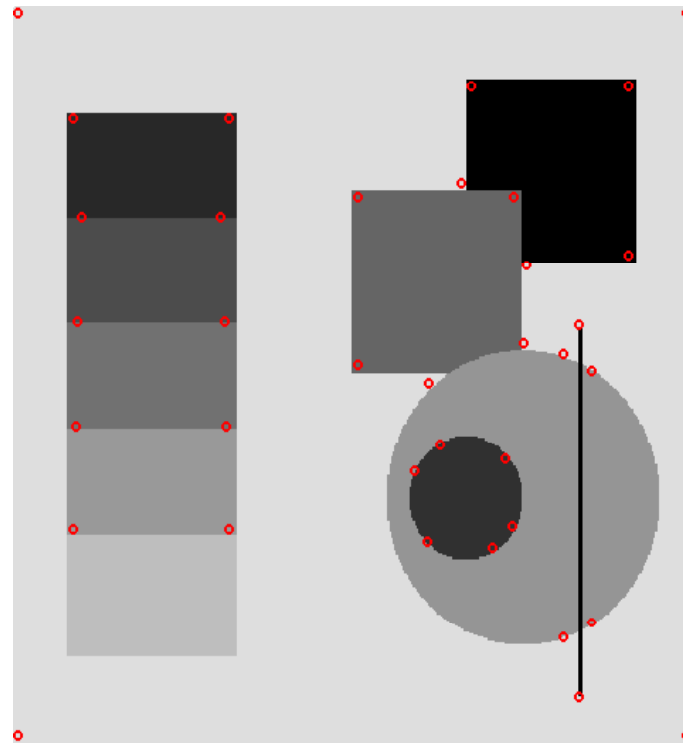
De az ábra feltételezi a kalibrált sztereó elrendezést

Legegyszerűbb illesztés (eddig) 4: Harris sarokpont detektor

A Harris operátor különösen érzékeny a skálázásra



1x



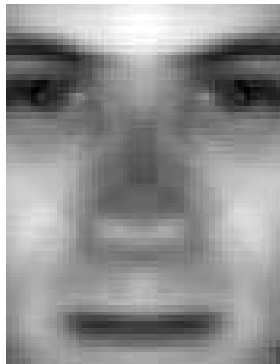
7x

Legegyszerűbb illesztés (eddig) 5

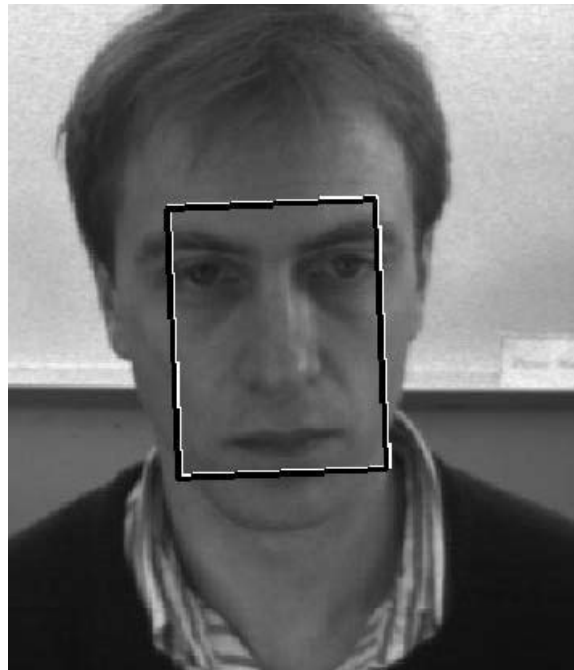
Textura követés

SSD (Sum-of-Squared-Differences) hiba minimalizálással keressük meg a legjobb illeszkedést

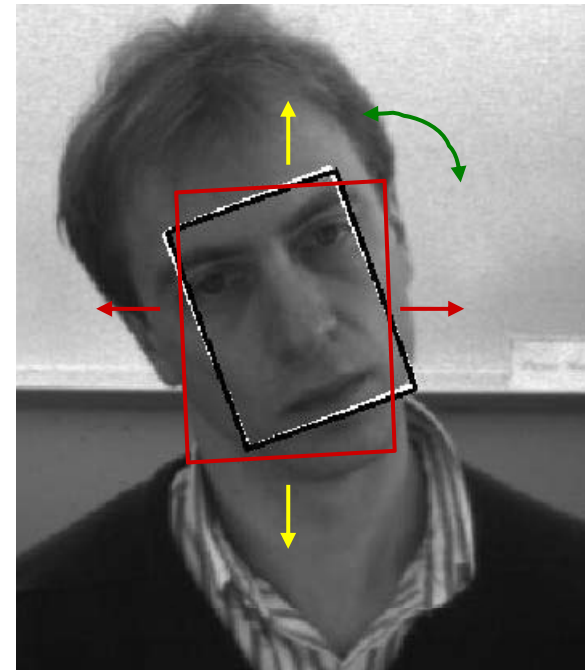
mintakép



A kezdő képkockán megtalálva



Piros keret: a kiinduló helyzet a következő képen



Esetünkben feltétlen valósídejűség kell: és a keresési idő mennyi?

Még robusztusabban detektálható és illeszthető releváns jellemzők (Interest Point)

Az elméleti rész viszonylag „friss” algoritmusokkal bővithető, például:

- KLT (1981-94) Kanade-Lucas-Tomasi feature tracker
- **SIFT** (1999) Scale Invariant Feature Transform
- SURF (2006) Speeded-Up Robust Features
-

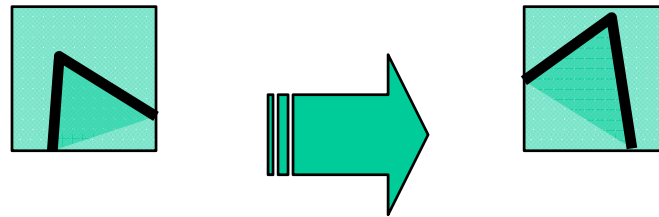
Jelentőségükhöz igazodóan számos platformra van nyitott forrású implementációjuk

Scale Invariant Feature Transform (SIFT)

A képfeldolgozás elméletének talán egyik legeredetibb, új (2000-es évek!) ötlete

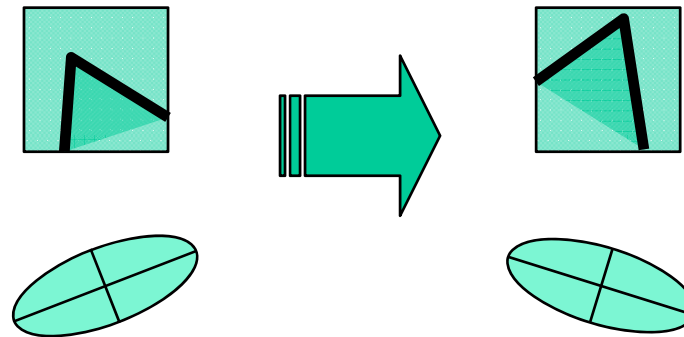
Ismétlés: Harris Detektor jellemzői

- Elforgatás invariancia?



Ismétlés: Harris Detektor jellemzői

- Elforgatás invariancia?



A „sajátvektor ellipszis” is forog de a sajátértékek változatlanok maradnak

R sarokság mérték tehát invariáns az elforgatásra

Ismétlés: Harris Detektor jellemzői

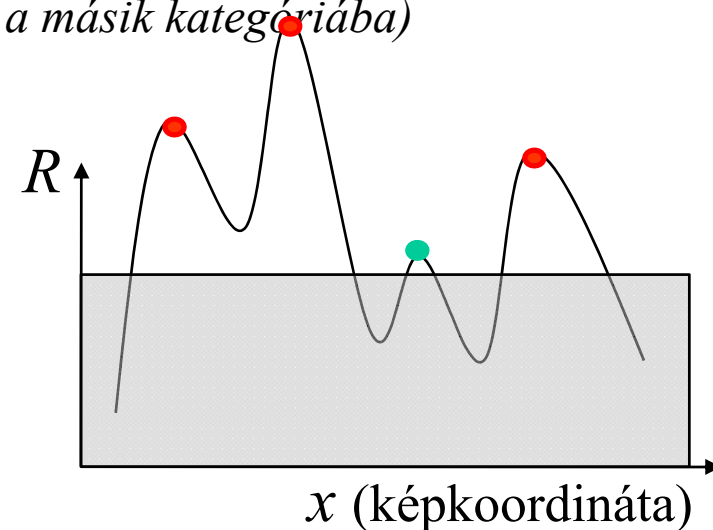
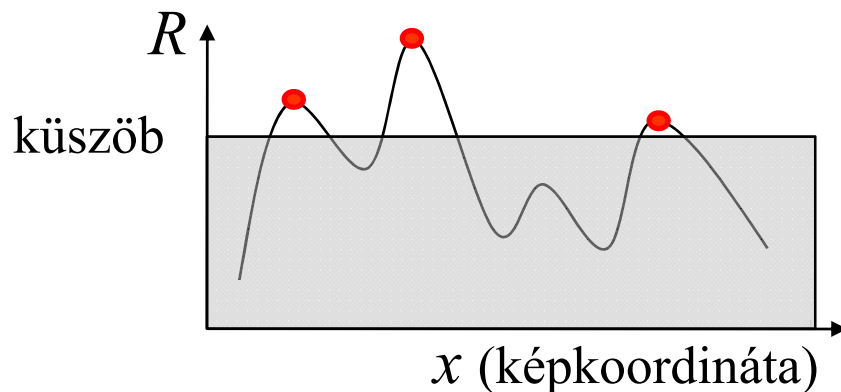
- Megvilágítás invariancia?

Ismétlés: Harris Detektor jellemzői

- Az additív és multiplikatív intenzitásváltozással szemben részleges az algoritmus invarianciája

✓ Intenzitás eltolás: *mivel a Harris csak deriváltakkal operál $\Rightarrow$ invariáns intenzitás eltolásra $I \rightarrow I + b$*

✓ Intenzitás skálázás: $I \rightarrow a I$ OK, kivéve az elégségesen nagy R értéket jellemző küszöb környékén (ha a nem a sarokság, hanem a multiplikatív intenzitásváltozás miatt kerül át a pont a másik kategóriába)



Ismétlés: Harris Detektor jellemzői

Skálázás invariancia?



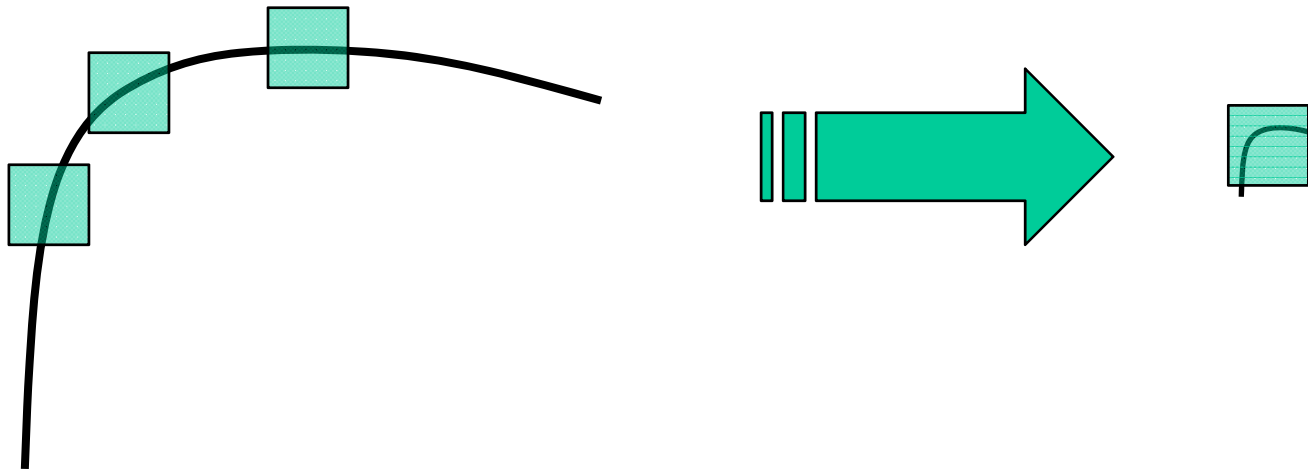
kép



Zoom-olt kép

Ismétlés: Harris Detektor jellemzői

- **Nem** invariáns a Harris a *képskálázásra*!



Minden pont élpontnak lesz
minősítve

Sarok !

Közbenső megjegyzések

- Sokat beszéltem eddig is a „robosztusan detektálható és követhető” jellemzőkről, de a követhetőség (egyediség) algoritmikusan kiszámítható mértékéről nagyvonalúan hallgattam (eddig)
- Először abban kellene biztosra mennünk, hogy ugyanazon „érdekes pontokat” (pl. sarokpontok) fogjuk mindegyik képen megtalálni, hogy a későbbi lépésben a követésre (párosításra) egyáltalán esélyünk legyen!
- Tehát, ha pl. sarokpontokat detektálunk – akkor azok legyenek relatíve kis számúak („nagyon sarkok”), és a geometriai skálázásra is tegyük valahogy érzéketlenné a detektálást!

Skálázás- invariáns függvény kell (egy példa)

http://www.robots.ox.ac.uk/~vgg/research/affine/det_eval_files/mikolajczyk_ijcv2004.pdf

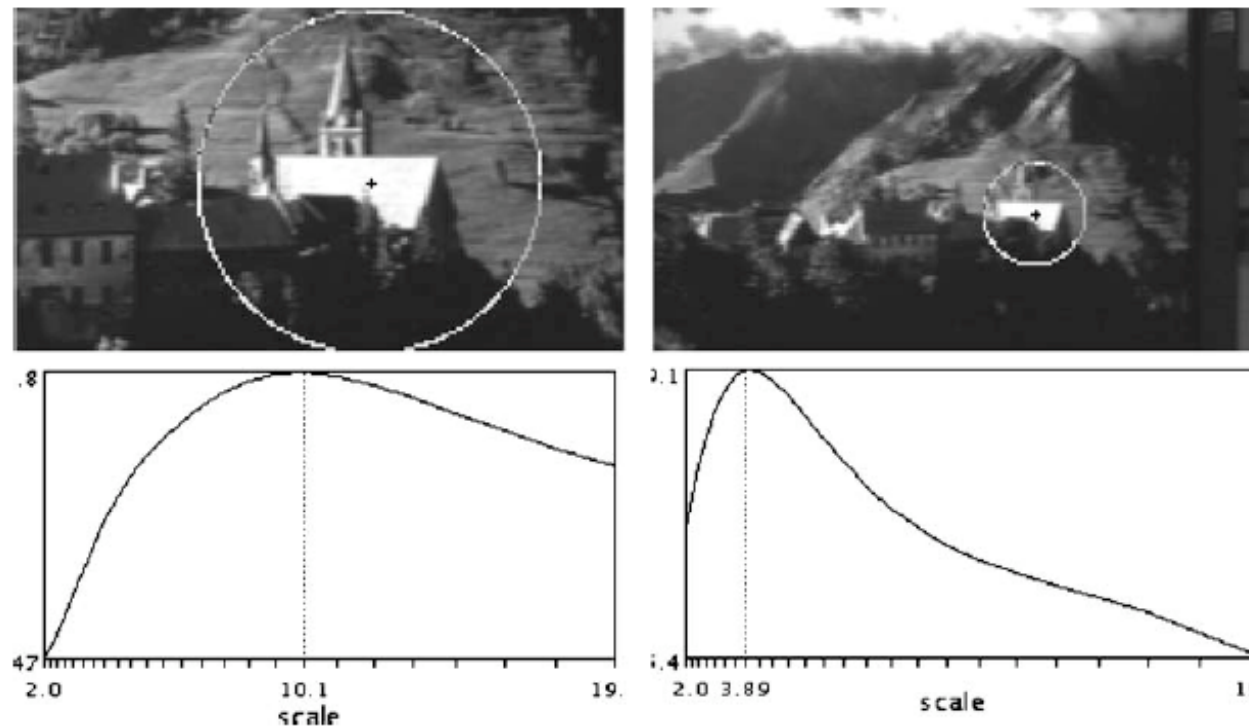


Figure 1. Example of characteristic scales. The top row shows two images taken with different focal lengths. The bottom row shows the response $F_{\text{norm}}(x, \sigma_n)$ over scales where F_{norm} is the normalized LoG (cf. Eq. (3)). The characteristic scales are 10.1 and 3.89 for the left and right image, respectively. The ratio of scales corresponds to the scale factor (2.5) between the two images. The radius of displayed regions in the top row is equal to 3 times the characteristic scale.

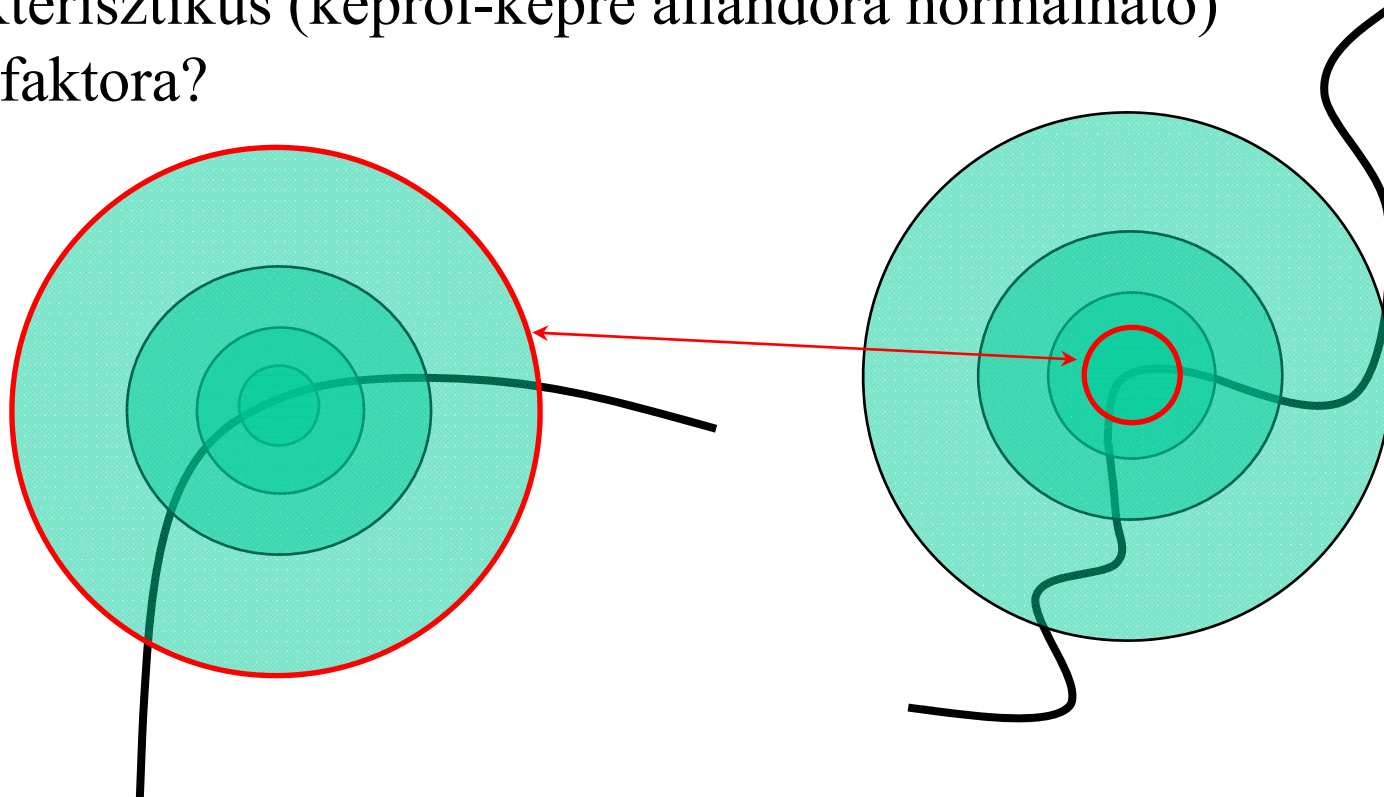
Skálázásra invariáns detektálás

- Különböző méretű régiókat (pl. köröket) nézzünk meg az érdekes pontok körül
- Az összehasonlítható régiókon belül hasonló képrészletet látnánk mindkét képen



Skálázásra invariáns detektálás

- A probléma: hogyan határozzuk meg az *összetartozó* köröket *függetlenül*, vagyis mindkét képen külön-külön is?
- Meghatározható az érdekes pontok környezetének egy karakterisztikus (képről-képre állandóra normálható) skálafaktora?



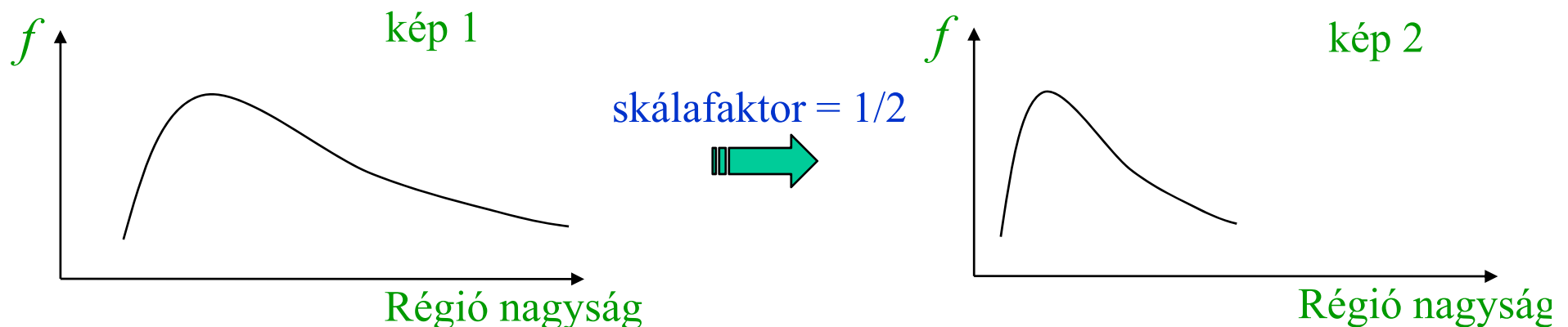
Ismétlés: Skálázás- invariáns detektálás

- **Az alapötlet:**

- Keressünk egy olyan függvényt, mely invariáns a körök átmérőjére (skálázás invariancia), vagyis a skálázástól függetlenül azonos a választ kapunk azonos képtartalomra

Egyszerű példa: **átlagos intenzitás** (azonos tartalmú régiókra nem változik a skálázással)

- A sugár függvényében ábrázolva a függvényt



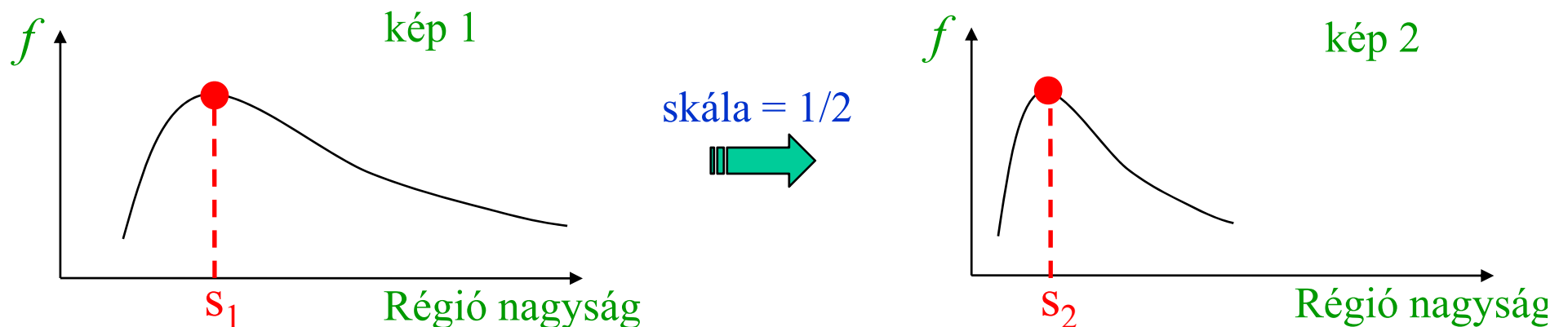
Skálázás-invariáns detektálás

- **Az alapötlet:**

Vegyünk a tetszőleges függvény **lokális maximumát** (a változó sugarú fv.-ben)

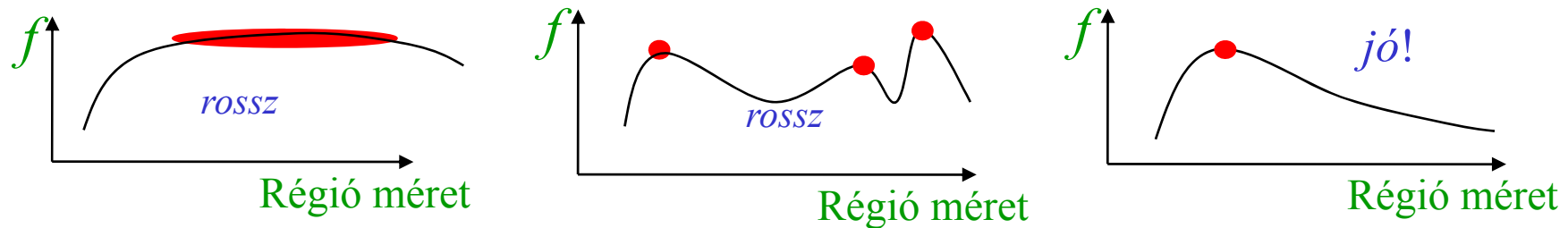
megfigyelés: az a régió nagyság, amelynél a maximum adódik **invariáns** a kép skálázásra, csakis az ablak alatti képtartalomtól függ!

Fontos: ez a skála-független régió méret minden képen - függetlenül a többi képtől- meghatározható!



Skálázás- invariáns detektálás

- Egy „jó” függvény kellene a skálázás detektálásra



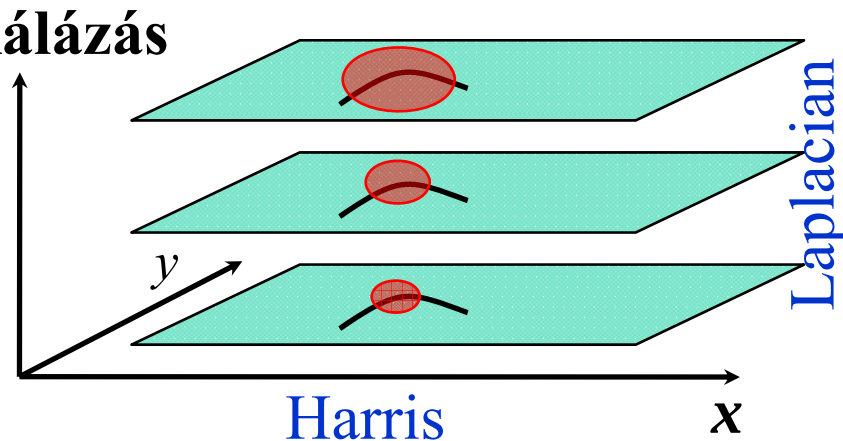
- Általában képekre igaz: az lesz egy „jó” függvény, amely kontrasztra (éles lokális intenzitás változásra) reagál

Skálázás- invariáns Harris detektor

Harris-Laplacian¹ skálázás

Keressük meg a lokális maximumokat az alábbi műveletekre:

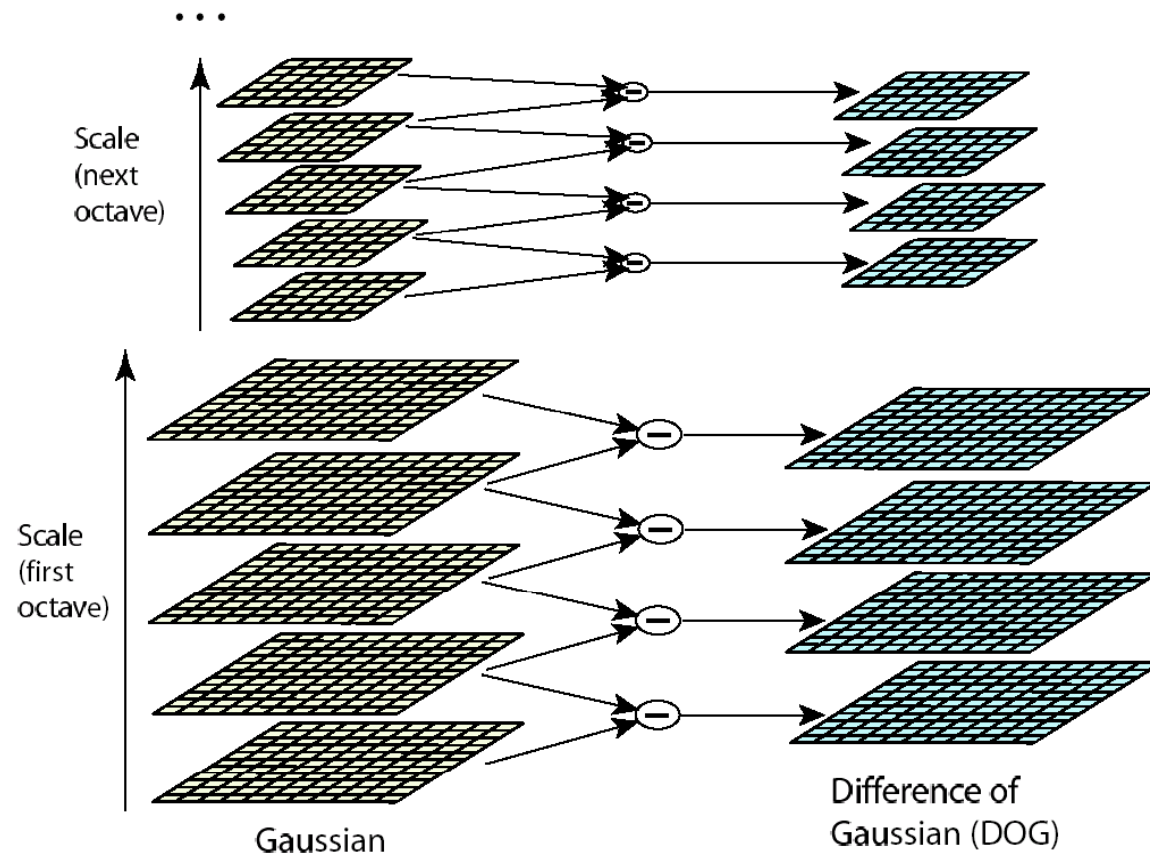
- Harris sarokpont operátorral az x, y képen
- Laplace operátorral a különböző újraskálázási faktor értékek mentén



¹ K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001

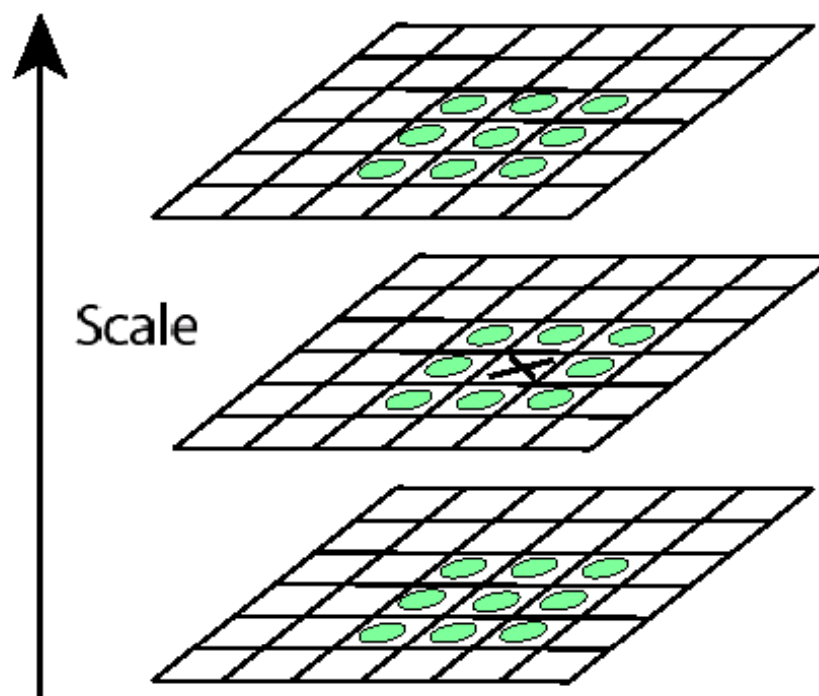
SIFT: a skálázás – képtér extrémum keresés algoritmus

Skálázás – képtér piramis építése

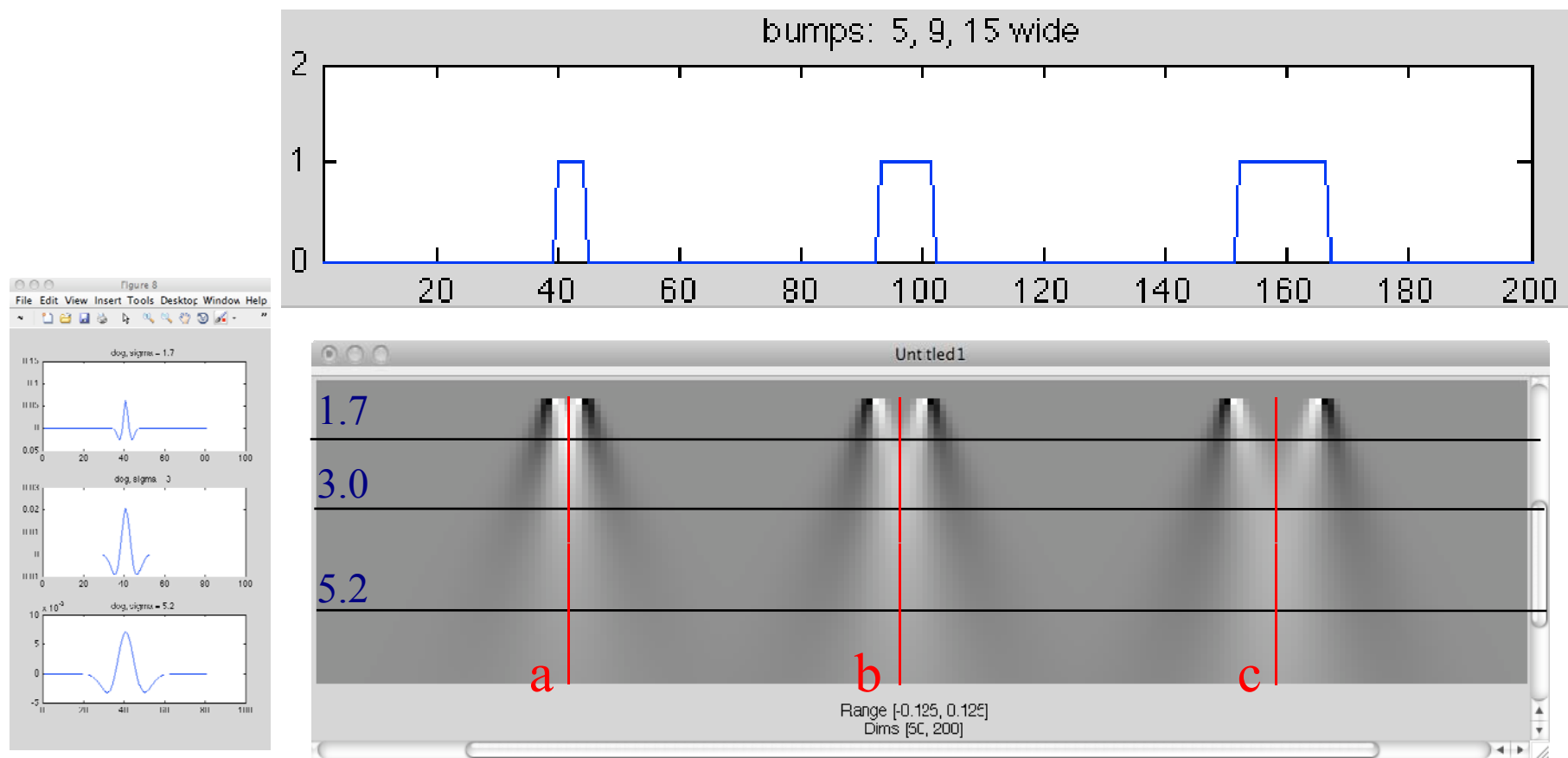


SIFT: a skálázás – képtér extrémum keresés algoritmus

A DOG függvényre a **képtérben és a skála mentén is** megkeressük a lokális maximumokat (egyenként az alatta és felette lévő skálaszinttel összevetve)



A csíkok Difference-of-Gaussian szűrővel megszűrve



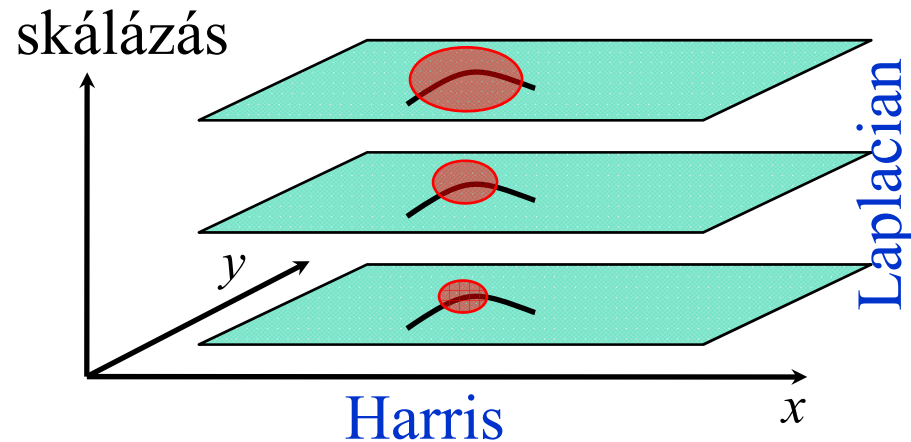
A következő dián megjelenített értékek pirossal

Harris-Laplacian vs. DoG (SIFT)

- **Harris-Laplacian**¹

Keressük a lokális maximumokat az alábbi műveletekre:

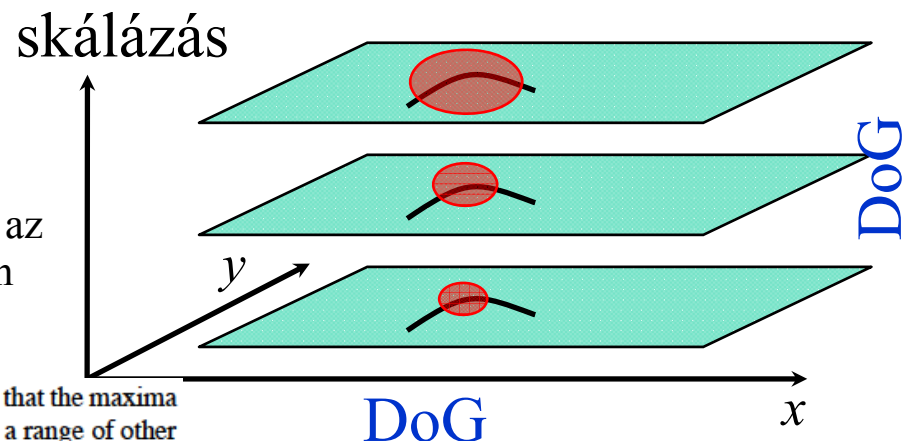
- Harris sarokpont operátor az x,y képen
- Laplace operátor különböző újraszkalázási faktor mentén



- **SIFT (Lowe)**²

Keressük a lokális maximumokat az alábbi műveletekre:

- DoG (Difference of Gaussians) operátor az x,y képen és az újraszkalázási faktor mentén



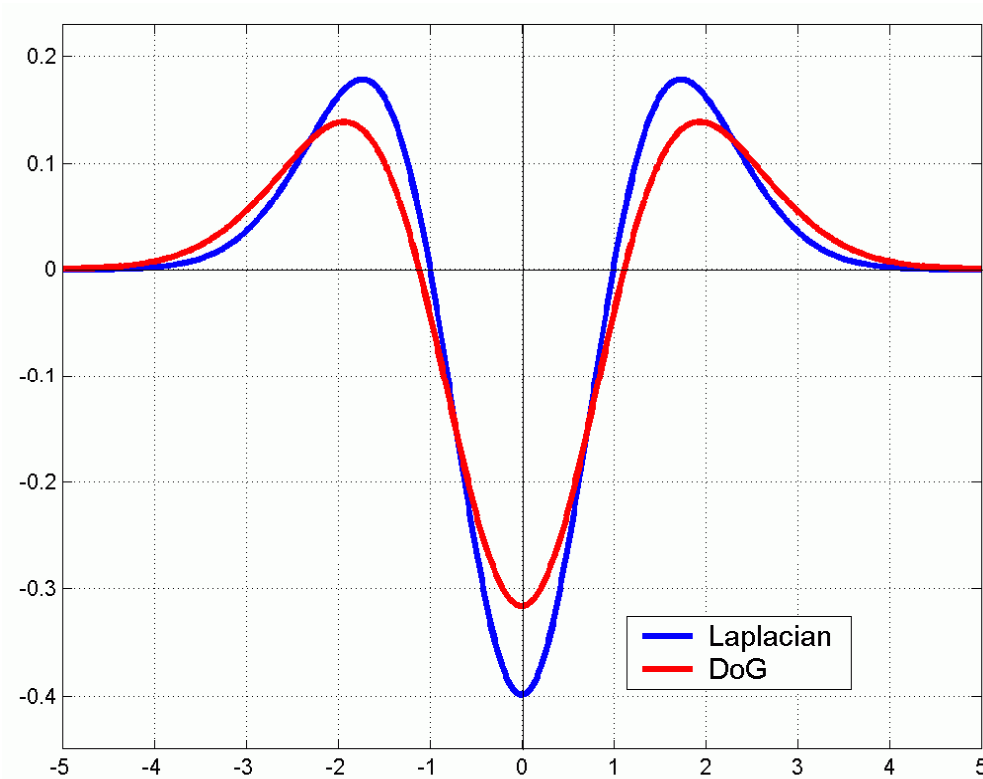
In detailed experimental comparisons, Mikolajczyk (2002) found that the maxima and minima of $\sigma^2 \nabla^2 G$ produce the most stable image features compared to a range of other possible image functions, such as the gradient, Hessian, or Harris corner function.

¹ K.Mikolajczyk, C.Schmid. “Indexing Based on Scale Invariant Interest Points”. ICCV 2001

² D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. Accepted to IJCV 2004

Laplace vs. DoG

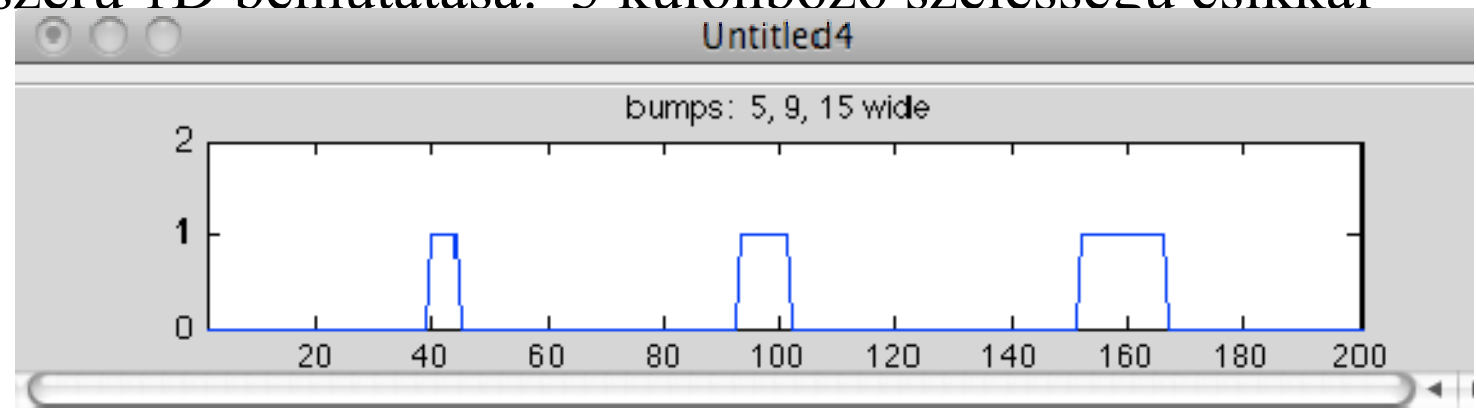
- Skálázási faktort jól normáló (mert attól arányosan függő relációban lévő) függvények összehasonlítása



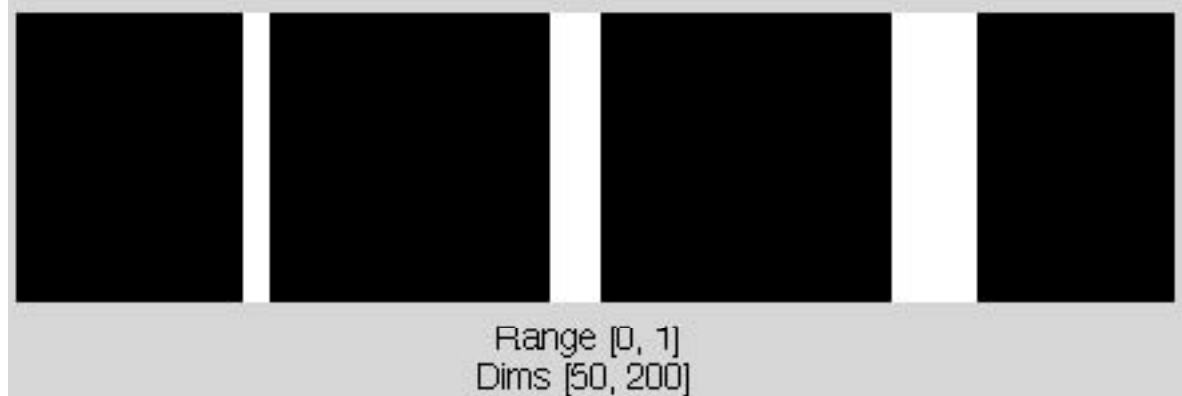
Példa: Skálázás- invariáns detektálásra

Az ötlet egyszerű 1D bemutatása: 3 különböző szélességű csíkkal

1D csíkok

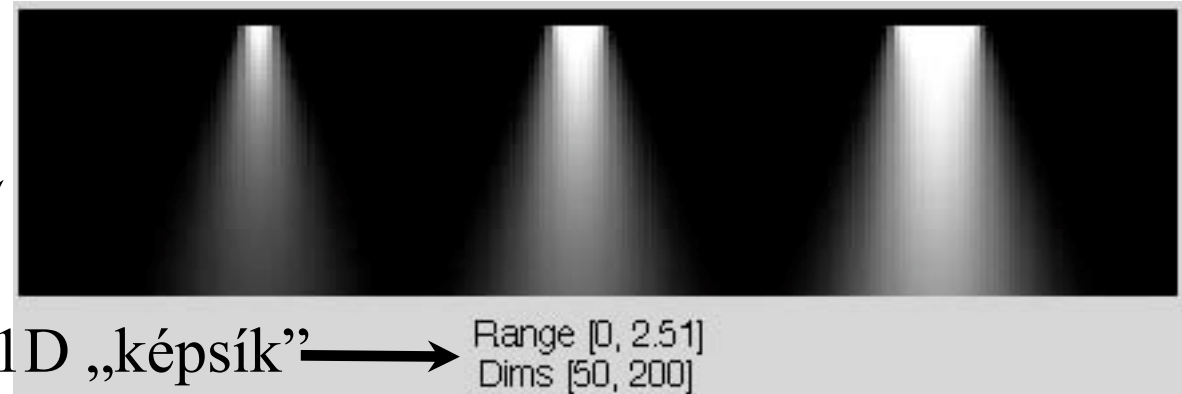
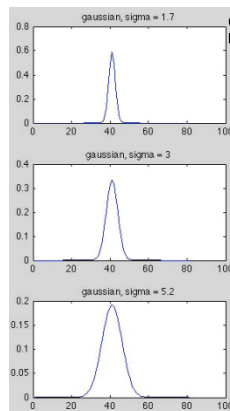


1D képi ábrázolás

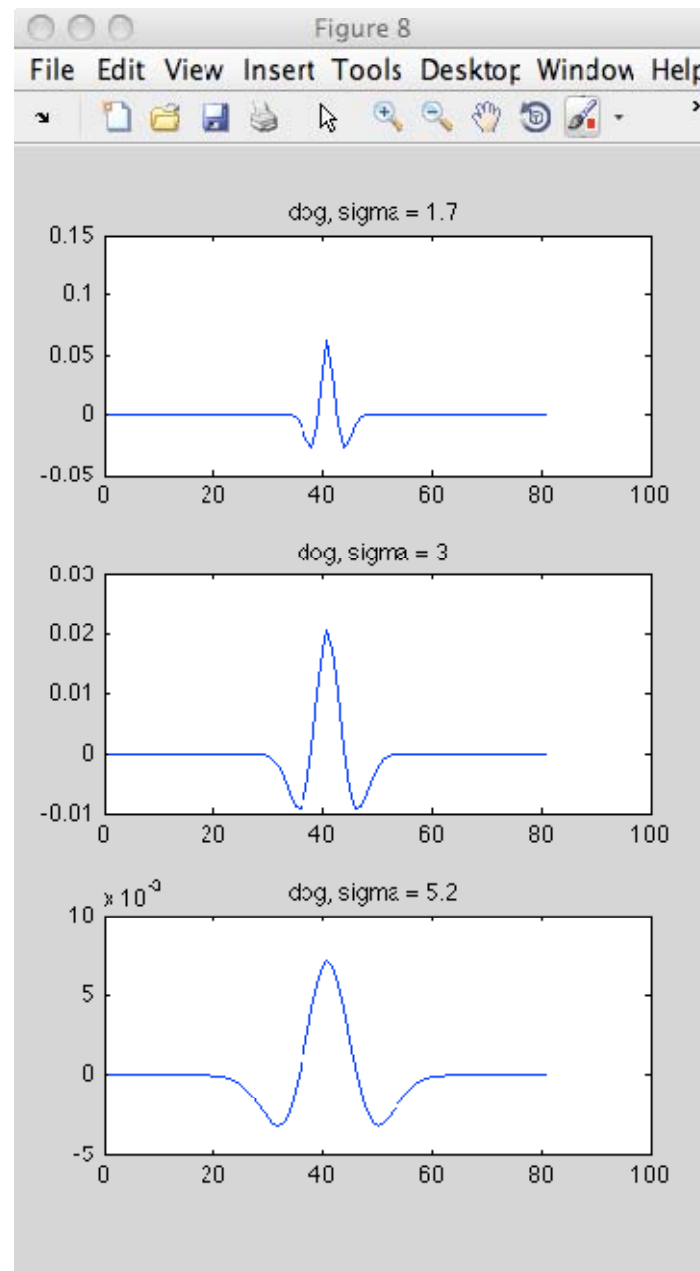


Elkenés
növekvő
szélességű
Gauss
szűrővel

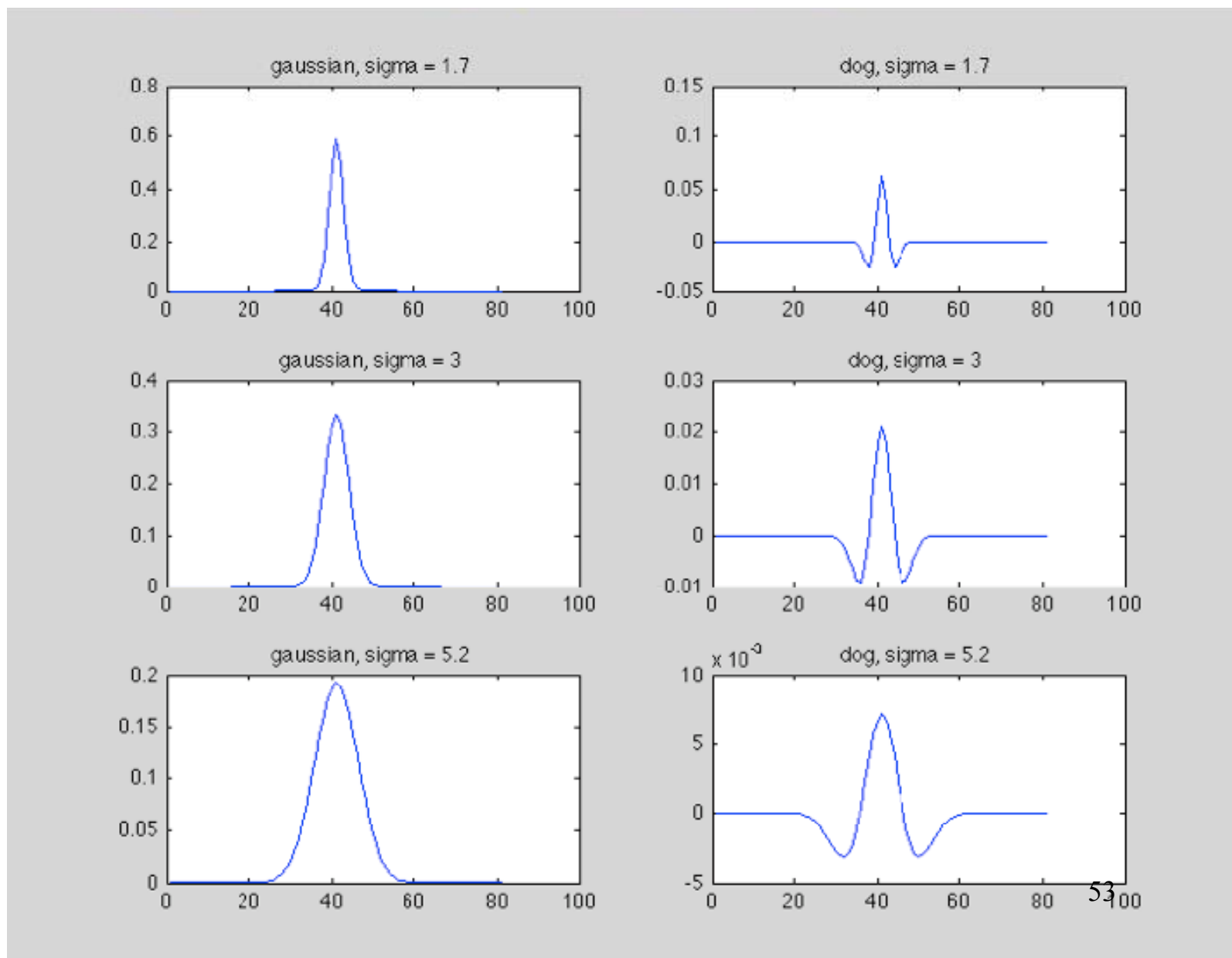
skálázás



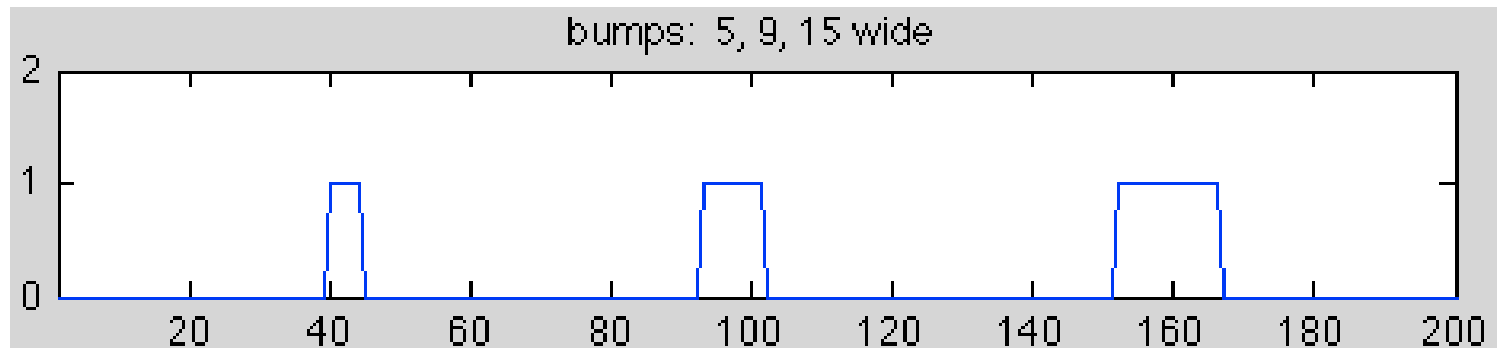
Skálázás- invariáns detektálás



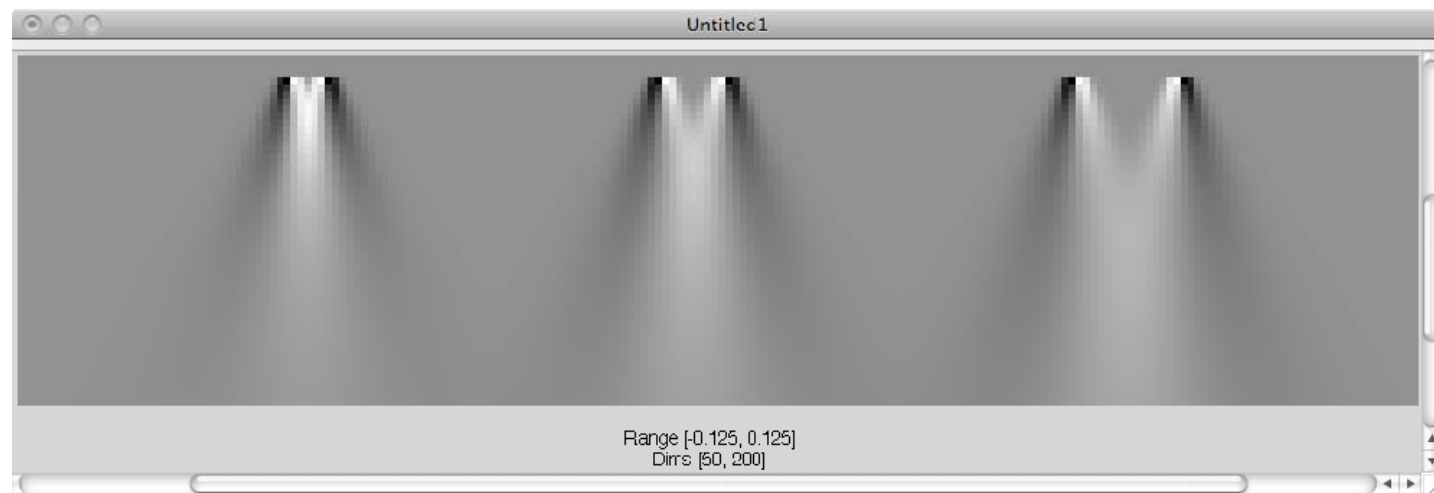
Gaussian és Difference-of-Gaussian szűrők



A csíkok Difference-of-Gaussian szűrve



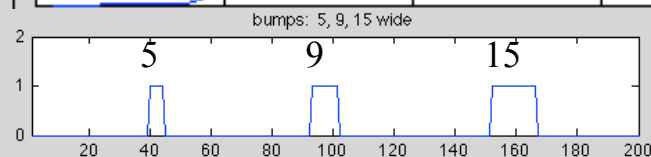
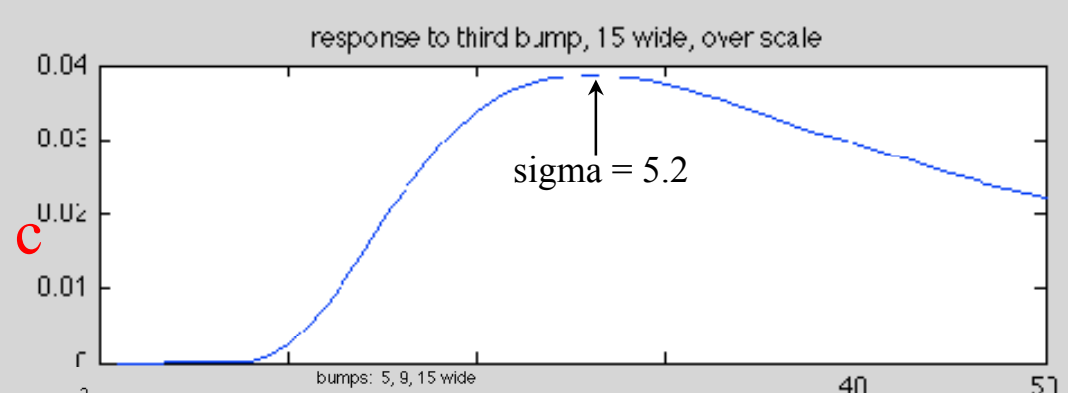
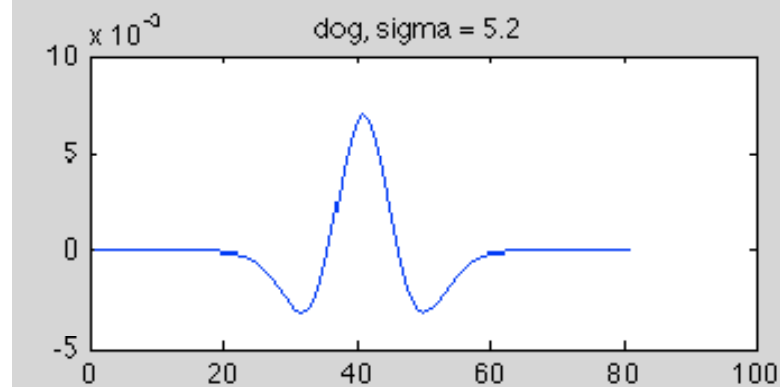
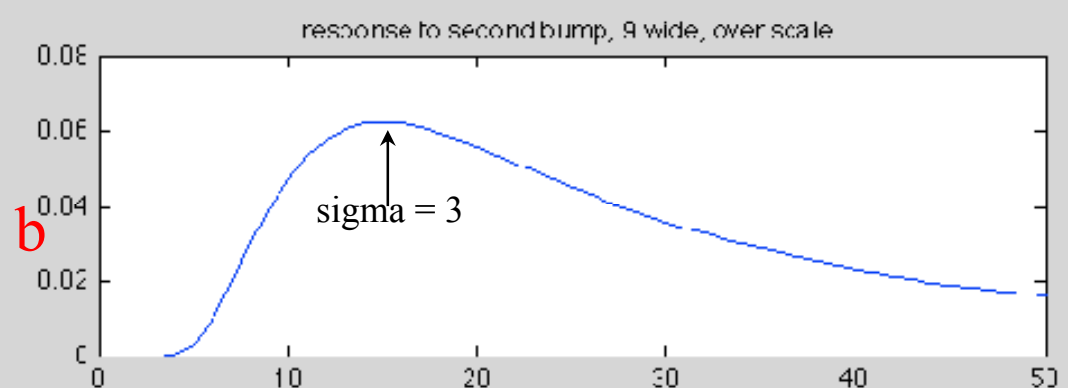
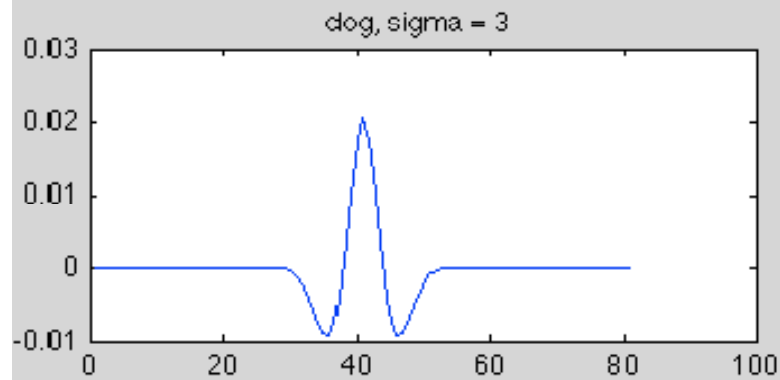
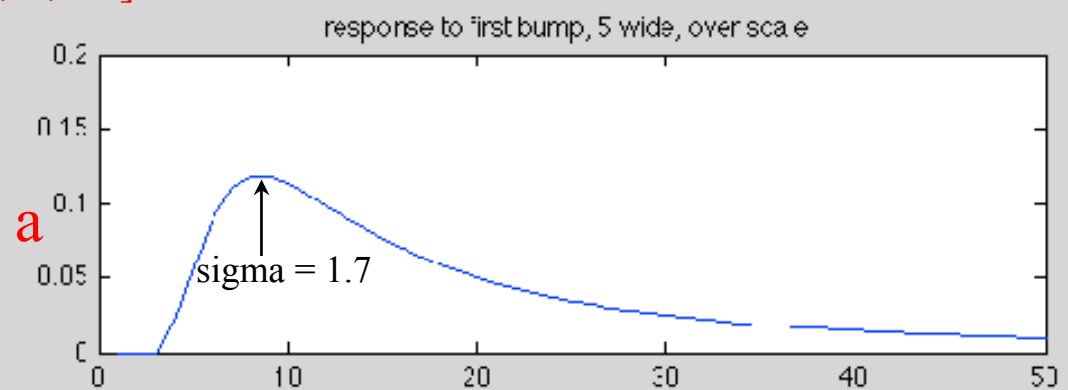
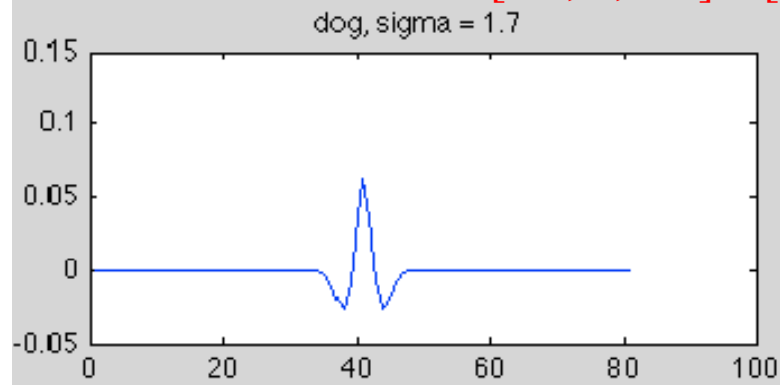
scale



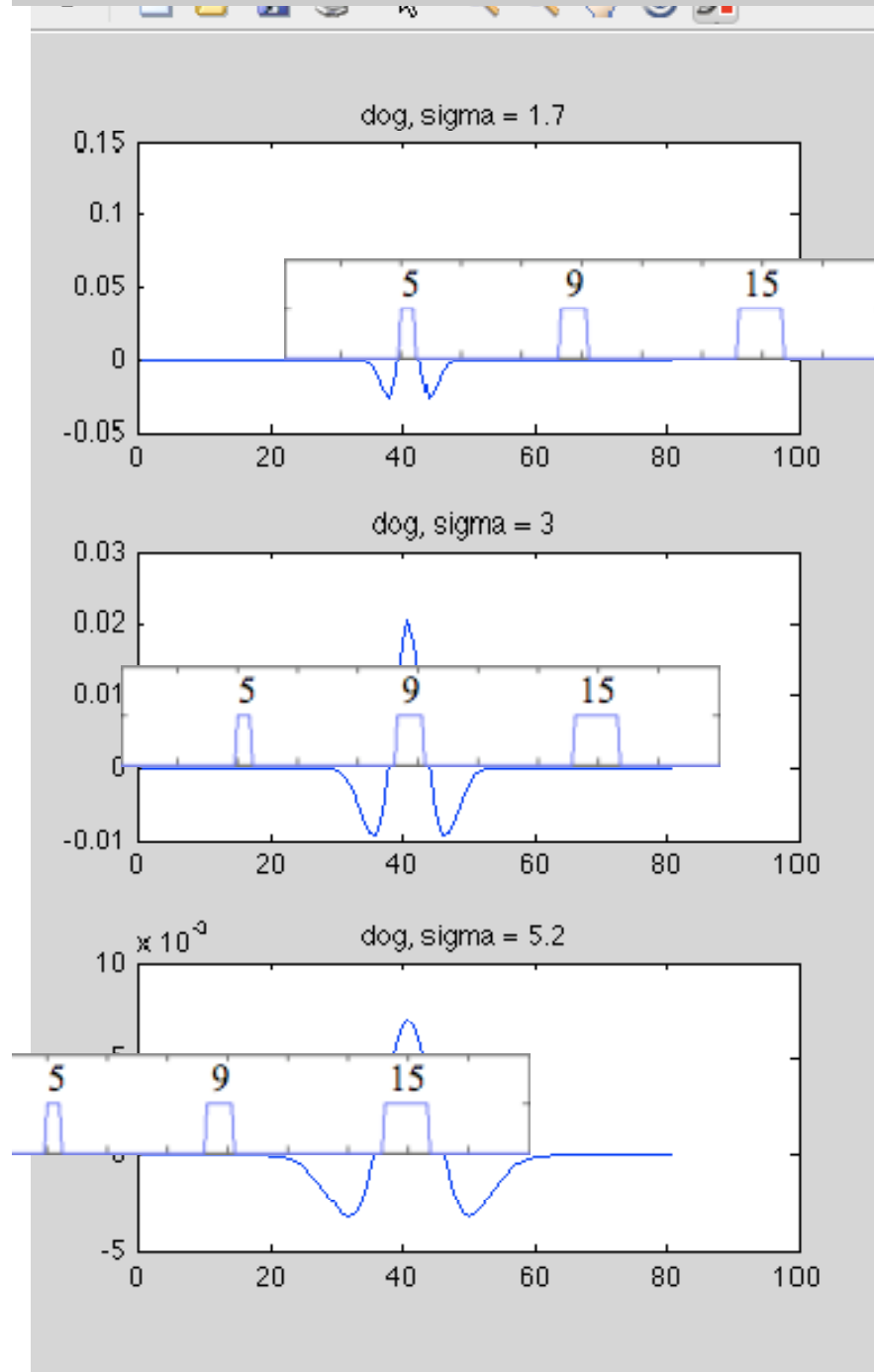
space →

A DoG szűrés utáni csúcsérték arányos a csíkszélességgel (csíkok különböző skálafaktor esetén):

$$[1.7, 3, 5.2] ./ [5, 9, 15] = 0.3400 \quad 0.3333 \quad 0.3467$$



Diff of Gauss szűrő válaszfüggvénye



$$[1.7, 3, 5.2] ./ [5, 9, 15] = 0.3400$$

$$0.3333 \quad 0.3467$$

Figyeljük meg:

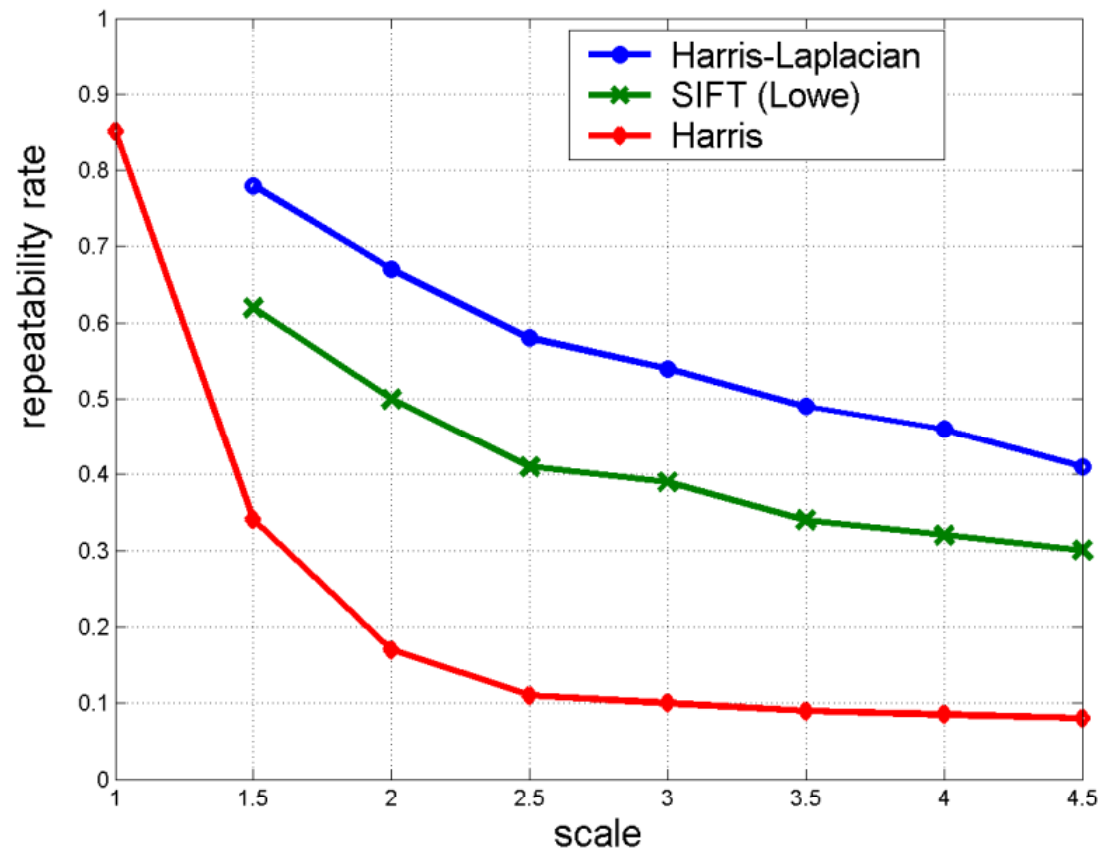
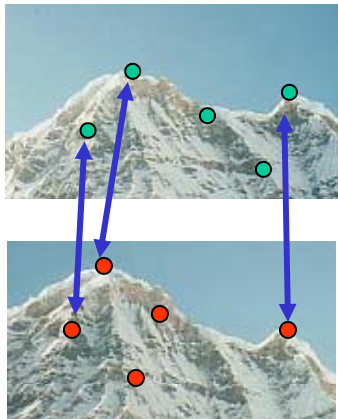
A maximális válaszú szűrő által „favorizált” csík minden felbontás mellett ugyanolyan relációban van az öt favorizáló szűrővel (a szűrő zéró-átmenetei mindig kb. az adott csík szélére esnek)

Vagyis a skála-kép analízis minden csíkból kiszedi a skálafüggést.

Ez a jelenség általánosítható a vizsgált 2D képekre is.

Skála invariáns detektorok összehasonlítása

„Repeatability rate” =
$$\frac{\text{megtalált illeszkedés}}{\text{keresett illeszkedés}}$$



Skála-invariáns jellemző detektálás (összefoglalás)

- adott két kép azonos valós látványról, de nagy skálázási eltéréssel
- **cél:** megtalálni az azonos „interest pontokat” (jellemzőket) mindkét képen
- **Megoldás:** egy megfelelően megválasztott függvény lokális maximumát megkeresni a kép és skálafaktor függvényében

Megismert konkrét módszerek:

1. **Harris-Laplacian** [Mikolajczyk, Schmid]: Laplace szerint lokális maximumot keresni a skála mentén + Harris szerint lokális maximumot keresni a képen
2. **SIFT** [Lowe]: Difference of Gaussians szerint lokális maximumot keresni a skála mentén és a képen egyaránt

Lokális Invariáns jellemzők előnyös tulajdonságai

- **Lokális jelleg:** ezért pl. kevésbé érzékeny takarásra,...
- **Megkülönböztethetőség:** segítségével nagyobb képi adatbázisban is eséllyel lehet keresni egyezőségeket
- **Mennyiség:** sok jellemző (interest point) kereshető
- **Hatékonyság:** pl. valósidejűség
- **Univerzalitás:** az alapelv könnyen kiterjeszthető a legkülönbözőbb jellemző típusokra – mindegyiket robusztusabbá téve

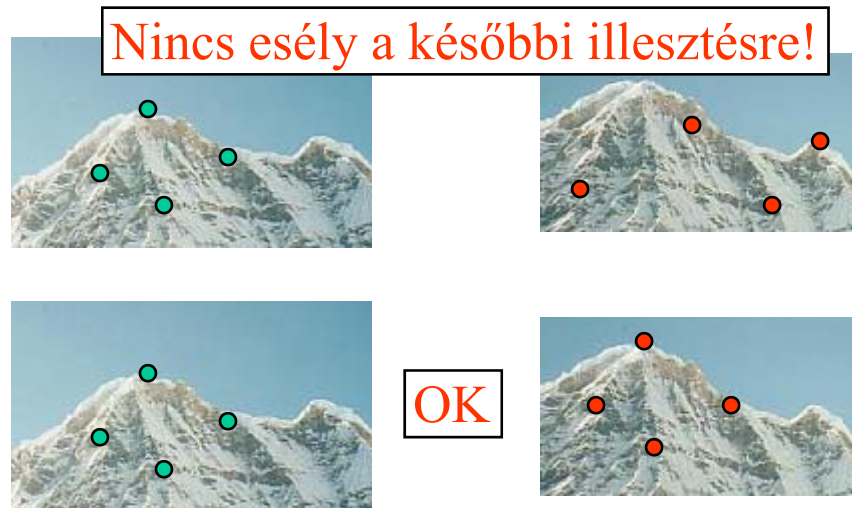
Hogyan tovább?



- Megvannak a skála- és rotáció-független (invariáns és megkülönböztethető) jellemzők („interest points”)
- Megvannak a skálázási faktorok
- Megvannak az orientációk
- A továbbiakban ezeket a pontokat kell úgy leírni (descriptor), hogy képről-képre beazonosíthatók legyenek

Emlékeztető: Képjellemzők

- 1. probléma:
 - Ugyanazon pontokat függetlenül megtalálni mindkét képen



Képről-képre reprodukálható
hatású **detektor** megtalálva!

Emlékeztető: Képjellemzők

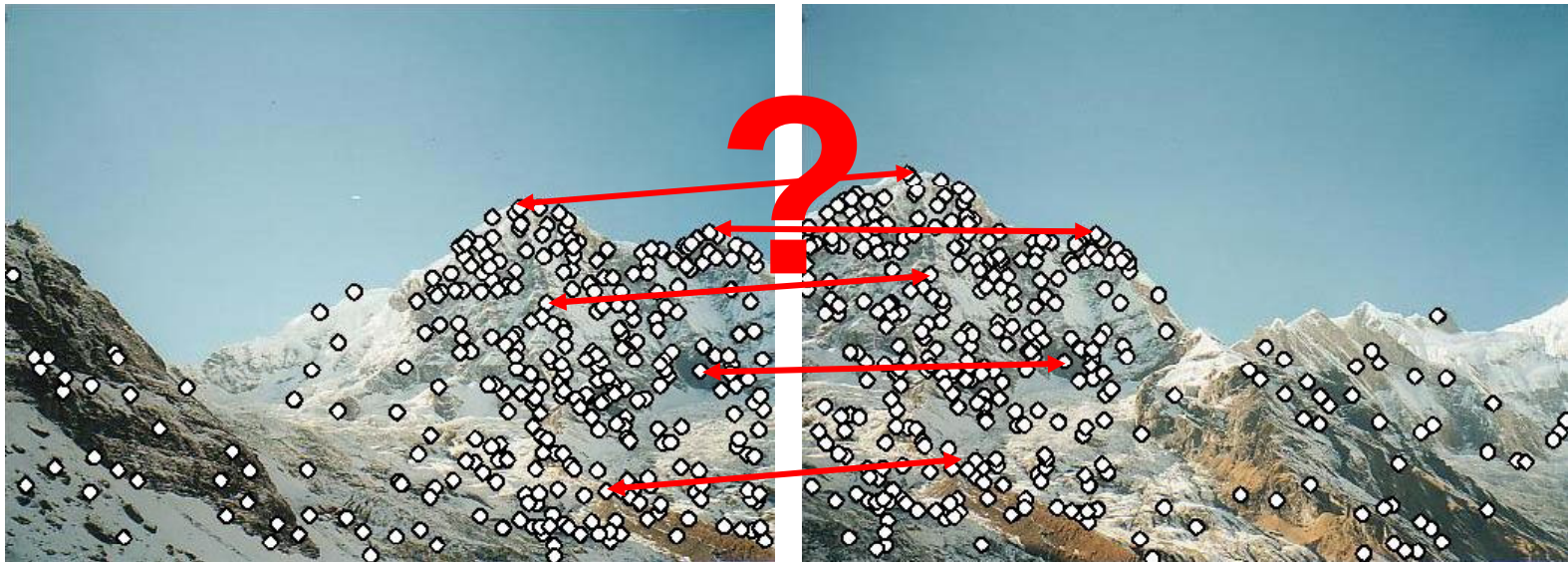
- 2. probléma:
 - Minden detektált ponthoz megtalálni az illeszkedőt a következő képen



Képről-képre megbízható és megkülönböztető
hatású **leíró** is kell még!

Megoldás: Jellemző pont leírása

A megtalált jellemzőket hogyan is illesszük?



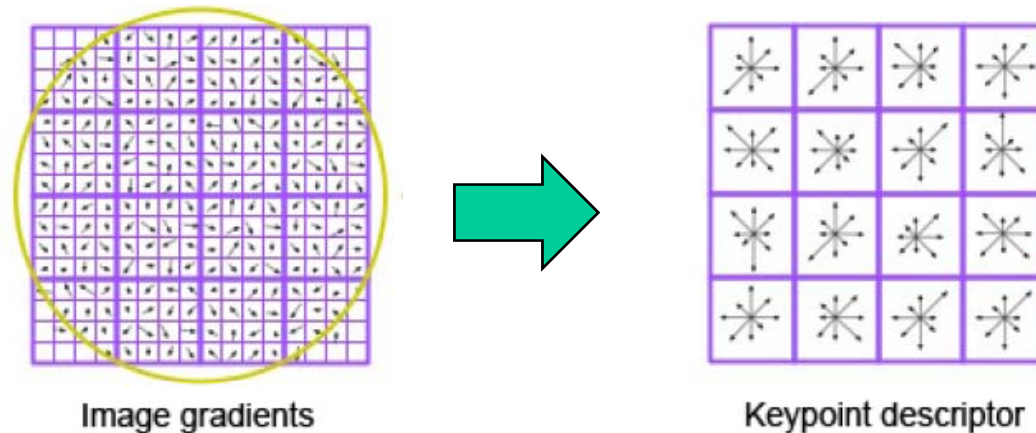
Pont leíró legyen:

1. Invariáns
2. Megkülönböztető (egyedi)

SIFT interest point leírás

1. Veszünk egy 16 x16 ablakot a „keypoint” körül
2. 4x4 cellákra bontjuk
3. Minden cellában kiszámoljuk a gradiensek hisztogramját

16 hisztogramm x 8 orientáció = 128 értékű jellemző vektorral keressük az illeszkedéseket

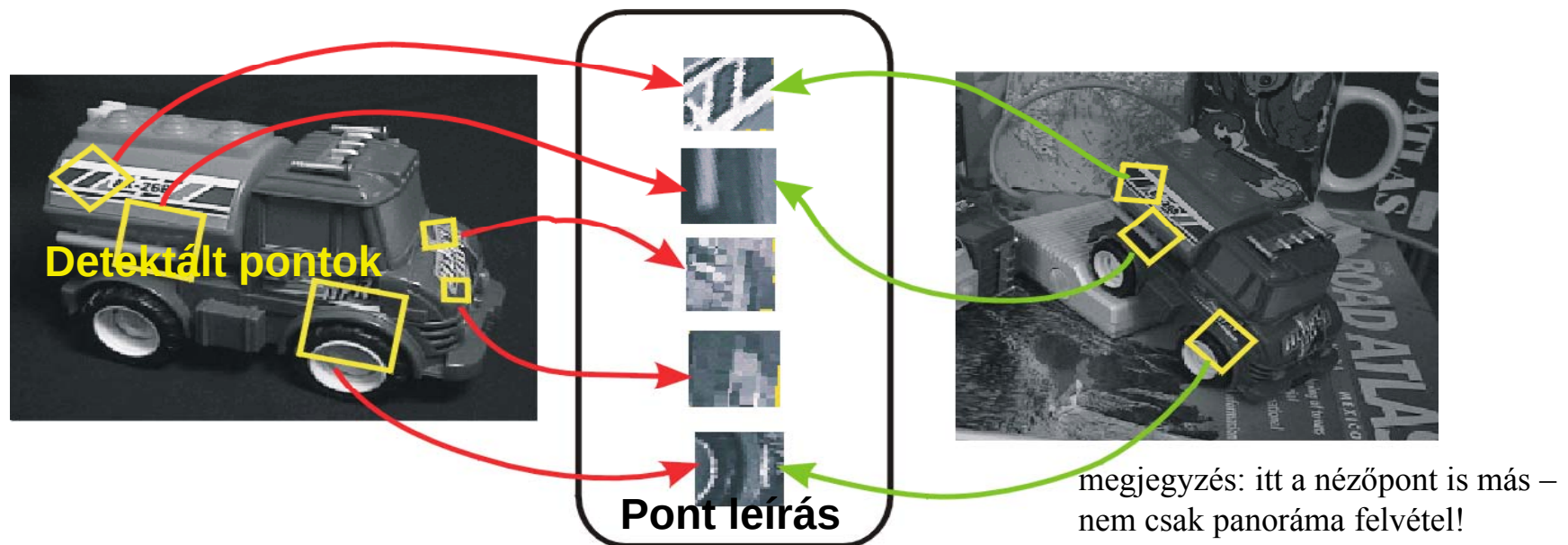


SIFT interest point leírás

- Megkülönböztető (jellemző környezetének egyediségét keresi):
 - Egyetlen jellemző pont is jó eséllyel korrektül illeszthető egy nagyobb adatbázisba (sok képen is megtalálva a „párját”).
- Skálázás és rotáció invariáns
- Továbbá: részlegesen invariáns
 - A kamera 3D-ben való nézőpontjának változására
 - Kb. 60 fokig az elforgatásra
 - Megvilágítás változásra
- Gyorsan számítható (real-time implementációk)

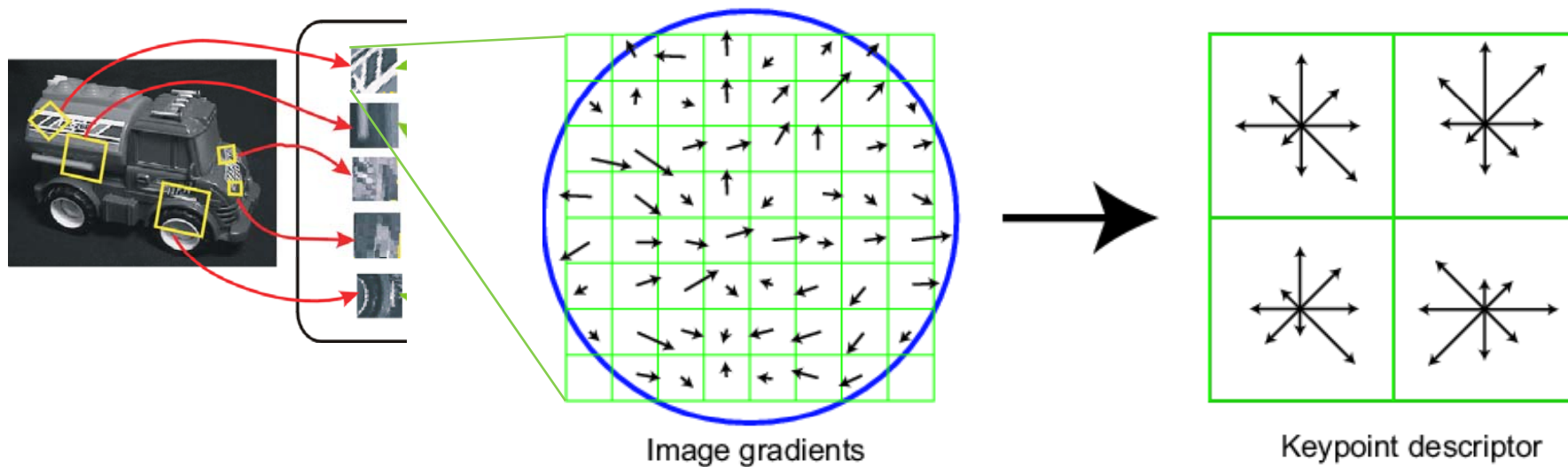
SIFT jellemző vektor generálás

1. **Detektálás:** ugyanazokat a jellemző pontokat nagy eséllyel megtalálja képről-képre (függetlenül a sorozatban lévő többi képkockától)
2. **Jellemző leírás:** egy lokális környezet szerint (már megint csak egy ablak vizsgálatára szorítkozunk!) skálázás & elforgatás független leírást rendel a jellemzőkhöz (többi jellemzőtől függetlenül kiszámítva)
3. **Illesztés:** megkeresni a leginkább „hasonló” (ld. különböző távolságmérték definíciók) leírókkal rendelkező jellemzőket a másik képen



SIFT jellemző vektor generálás

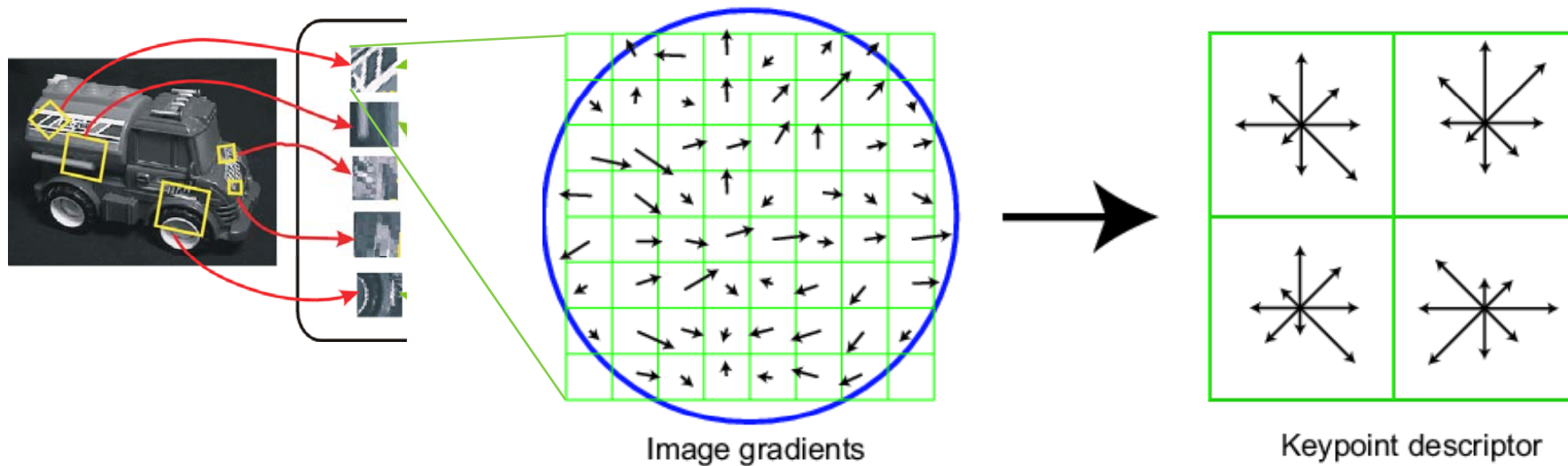
- Egy 4x4 mezőben küszöböztött orientációkat számolunk a gradiensek nagyságával súlyozva
- 8 lehetséges orientáció x 4x4 mező = 128 dimenziós jellemző leírás



Csak egy 2x2 részlet bemutatva a 4x4 mezőből

SIFT - Intenzitásváltozás invariancia biztosítása

- 128-dimenziós vektor normálásával (normálás 1-re)

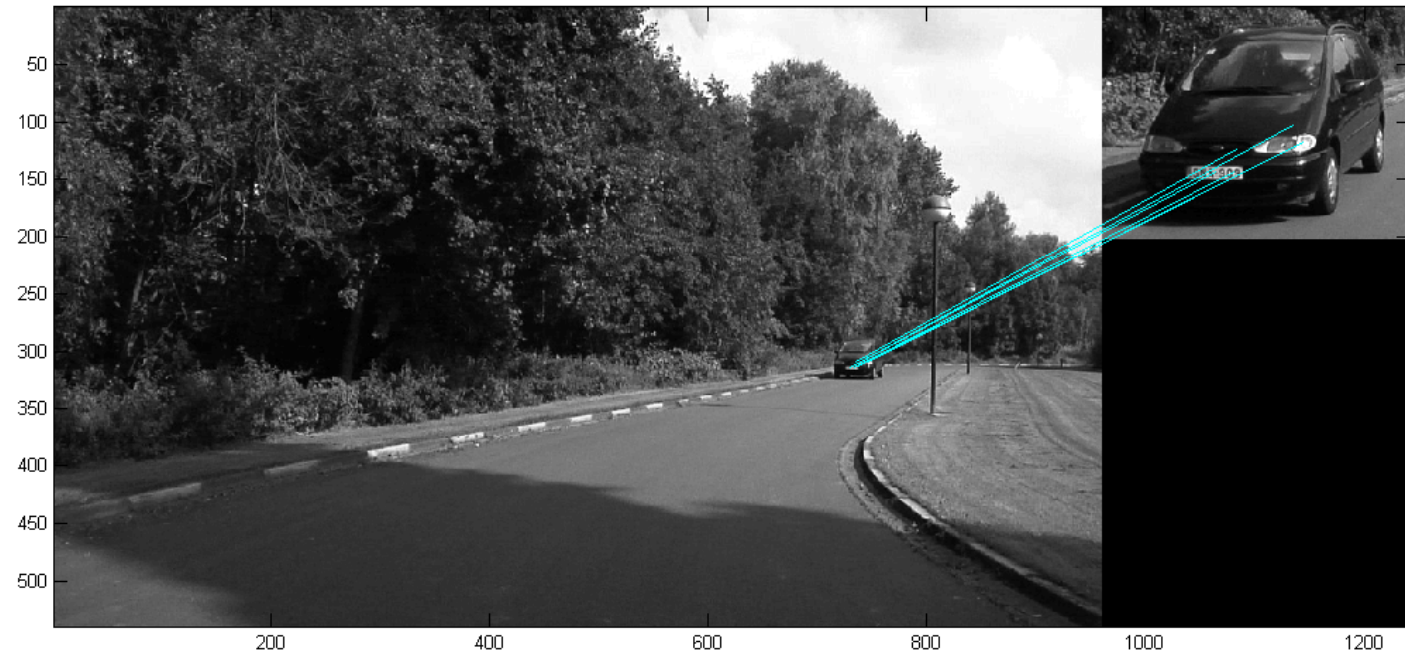


Mit is kaptunk végeredményben?

- Képről-képre jól lokalizálható jellemzőket (interest points)
- Ezeket a jellemzőket (interest points) jól megkülönböztető leírásokat (128 dimenziós vektorokkal)
- Ami ezek után már megtehető:
 - Az egyes képeken függetlenül megtalált invariáns jellemző pontok illesztése képről képre a hozzájuk rendelt leírás alapján
 - Azonosság (homography) robusztus meghatározása – dacára a torzításoknak, takarásoknak, stb.

Demo – recognition of a car

$t = 0$ ms



Five points from the template are correctly identified in the scene

Demo – recognition of a car

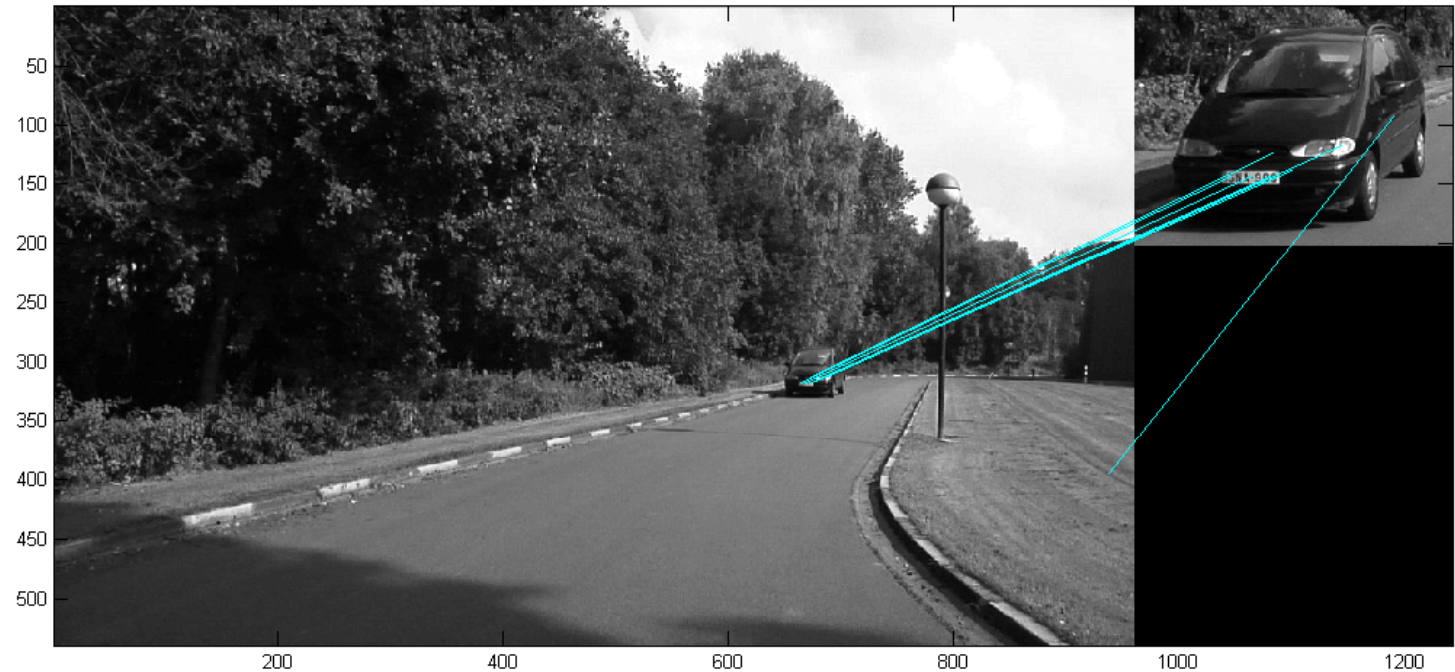
We will use a template of a car and try to match it against a scene in which this car is present



template

Demo – recognition of a car

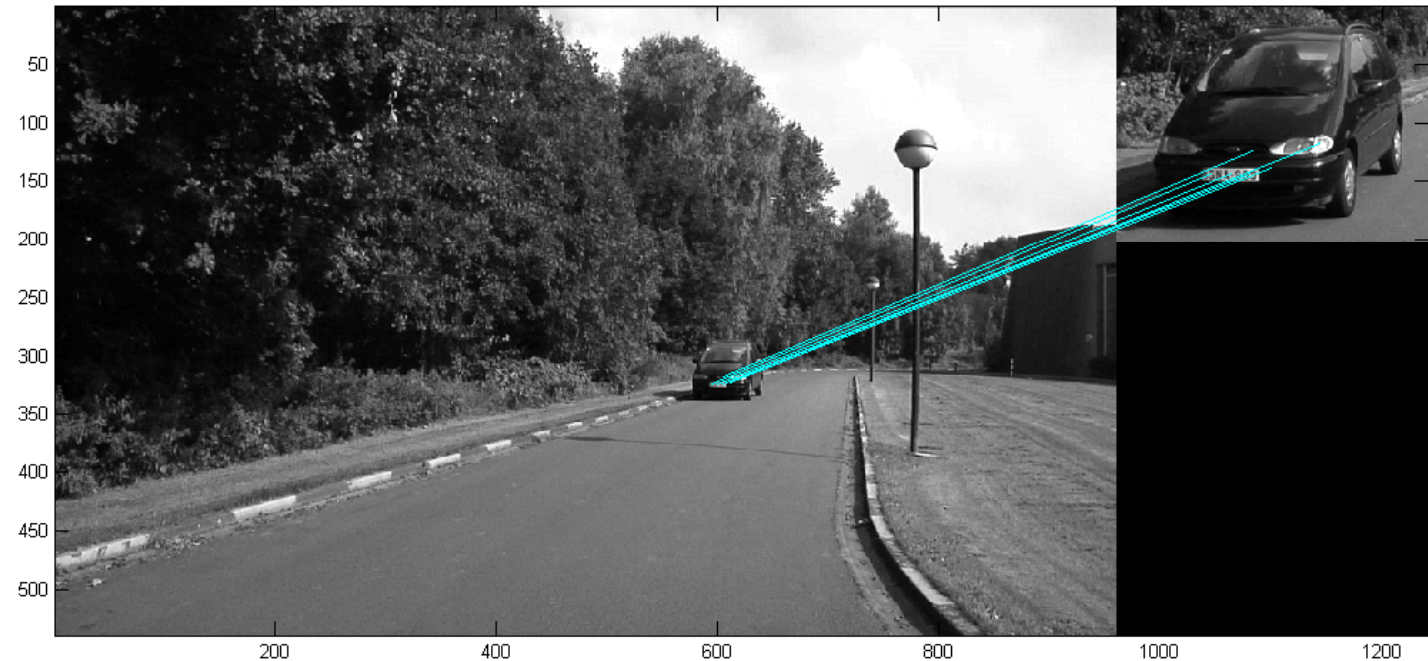
$t = 400$ ms



Six points from the template are correctly identified in the scene, however also note the incorrect match in the right of the image

Demo – recognition of a car

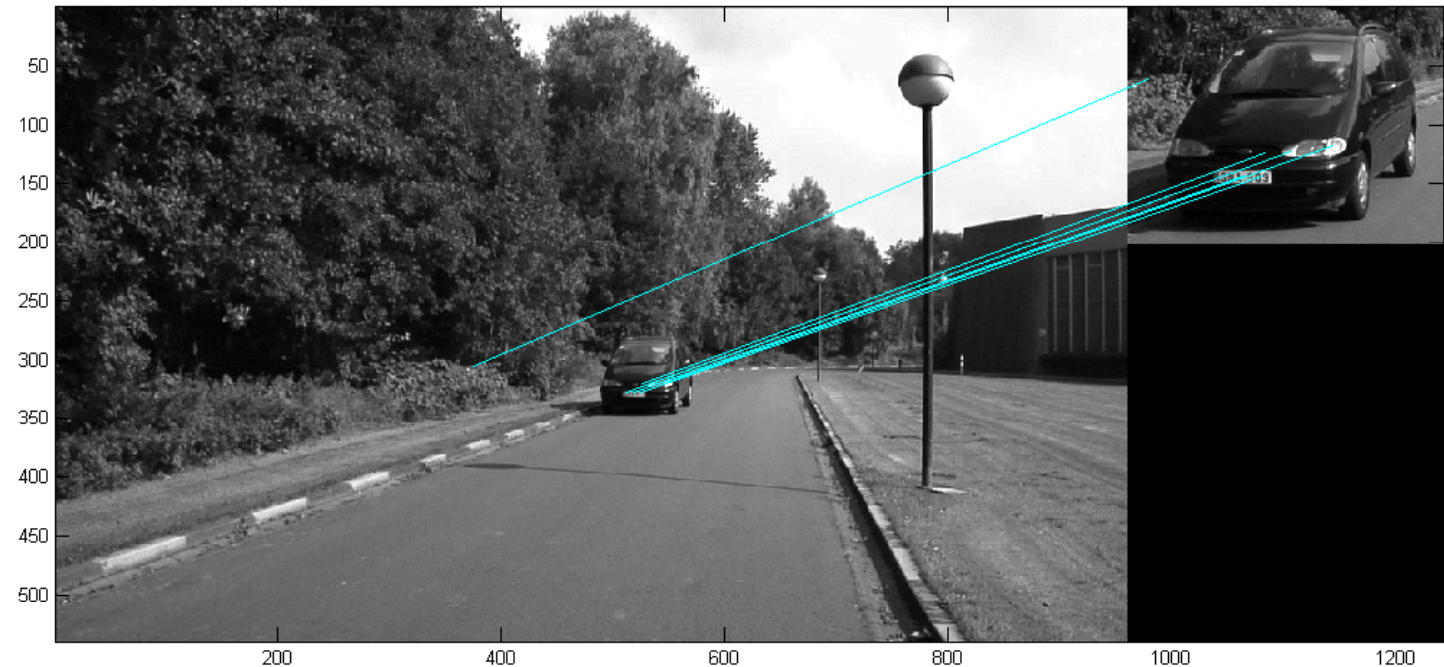
$t = 800$ ms



Six points from the template are correctly identified in the scene.

Demo – recognition of a car

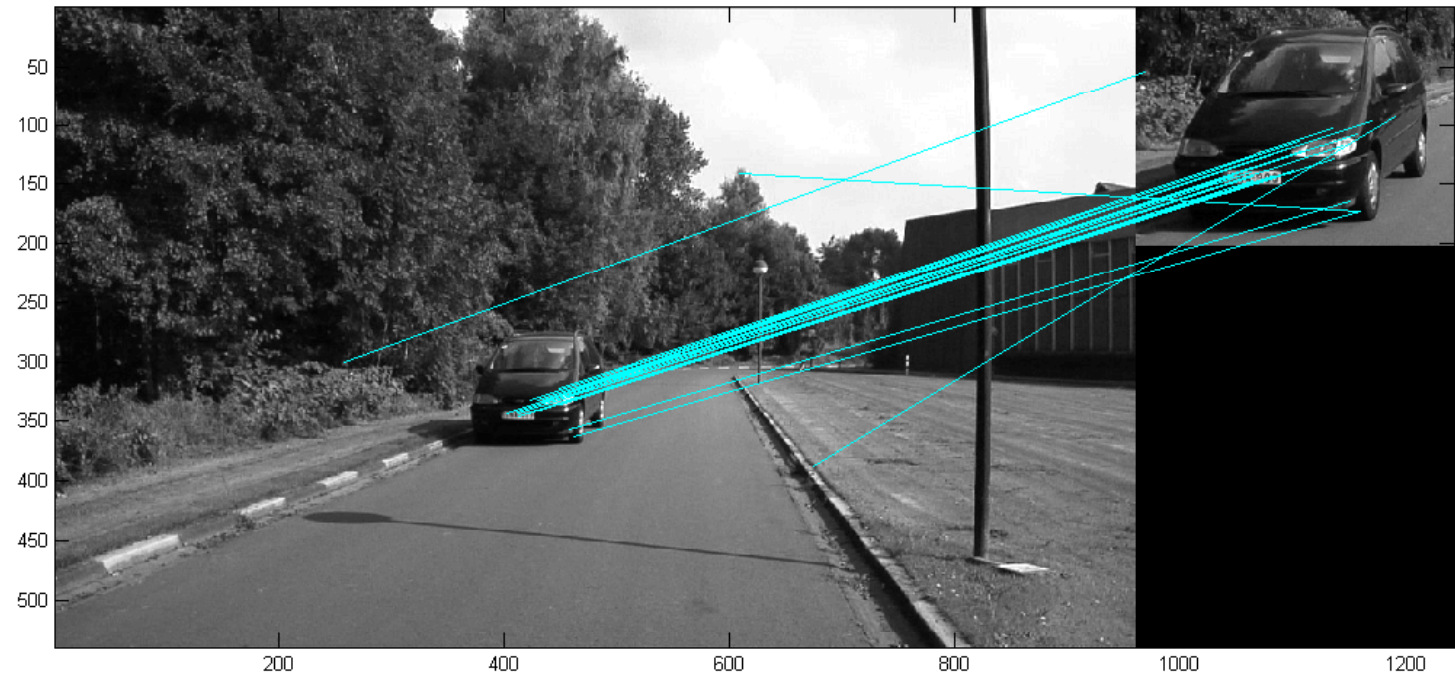
$t = 1200 \text{ ms}$



six points from the template are correctly identified in the scene (one point does not belong to the car however).

Demo – recognition of a car

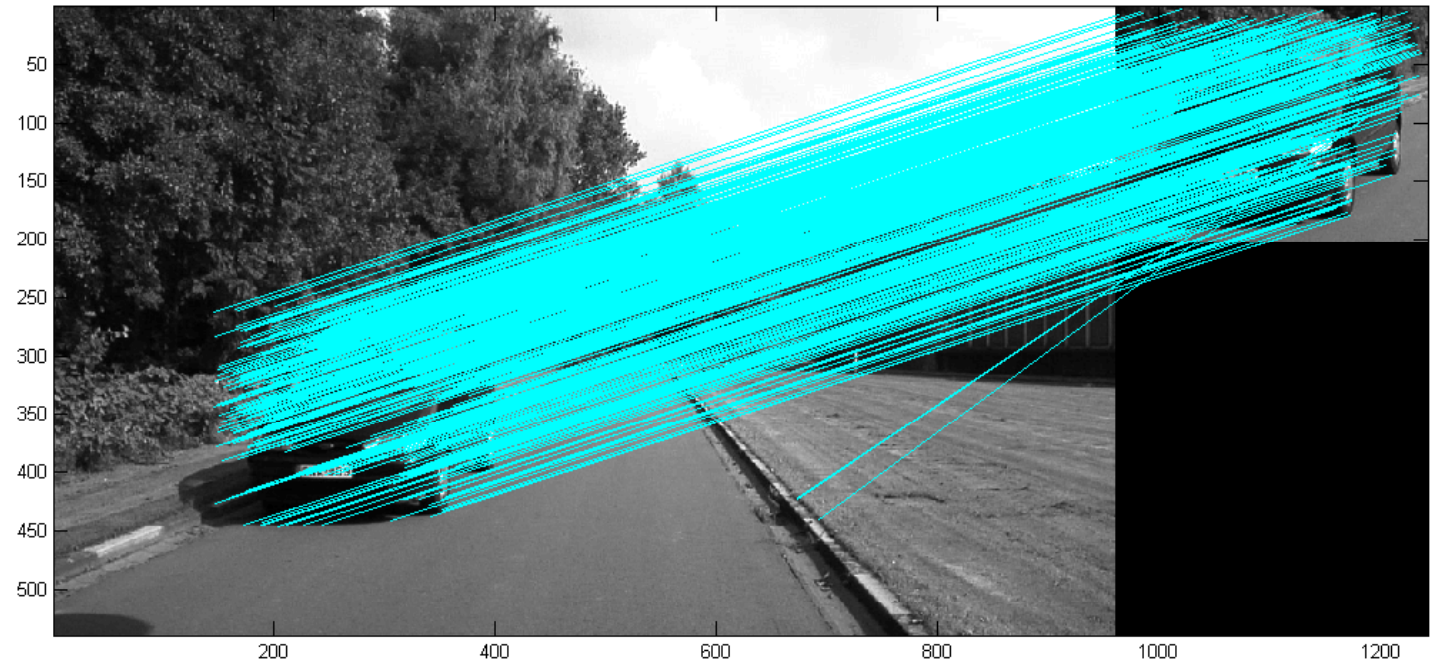
$t = 1600$ ms



More points are being recognized, two points are wrongly matched

Demo – recognition of a car

$t = 2000$ ms



A lot of points are correctly matched (this is to be expected since the template was derived from this image). Two points are incorrectly matched

Lokális ablakműveletek

- Idő hiányában SIFT-el abbahagyjuk...
- Pedig lehetne tovább építkezni, csupán a valós időben jól implementálható lokális műveletekre szorítkozva...

Az eddigi lokális ablakműveletek rész
néhány általános tanulsága

A jó képfeldolgozási stratégia

Az optimális képfeldolgozási útvonalak kiválasztásának és algoritmikus bejárásának képességét kell elsősorban megszerezni!

A stratégia két lépésben összefoglalva:

1. a megvalósítandó funkciókhoz megtalálni a „legjobb” algoritmust
2. az így megvalósítandó algoritmushoz megtalálni a „legjobb” architektúrát.