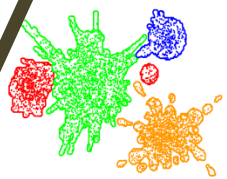
Recunoasterea de gesturi
Urmarirea obiectelor

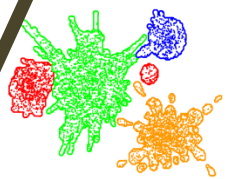
Interfata vizuala om-masina

Laura Florea



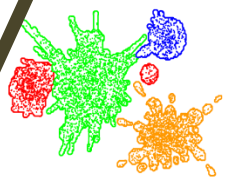
Cuprins

1. Problema recunoasterii de gesturi
2. Localizarea punctelor importante in imagini
3. Metode de urmarire a obiectelor



Problema recunoasterii de gesturi
Introducere

Interfata vizuala om-masina



Aplicatii









Medicale:

- Ajutor pentru persoanele cu dificultăți de auz, precum recunoașterea automată a limbajului semnelor
- Chirurgie
- Monitorizare la nivel emotional

Criminalistica:

- Tehnici de identificare
- Detector de minciuni

Securitate:

- Detectie de comportament anormal

Aplicatii



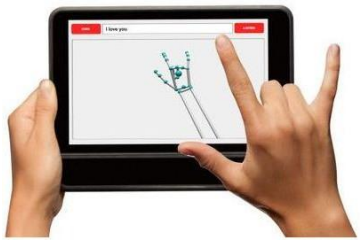
Divertisment:

- “gaming”;
- navigare și manipulare în medii virtuale și de “realitate augmentată”.



Industrie:

- manipulare echipamente / roboţi industriali.



Asistare persoane cu dizabilități:

- ex: tabletă cu recunoaștere automată a semnelor.

Noțiuni introductive

- Tendinta actuala – inlocuire a dispozitivelor mecanice clasice





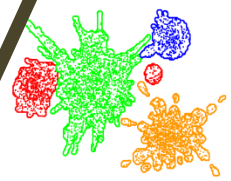








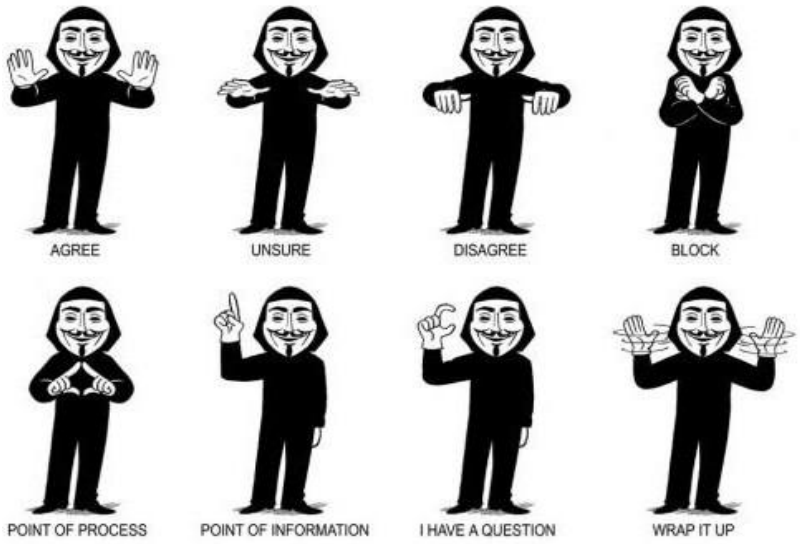




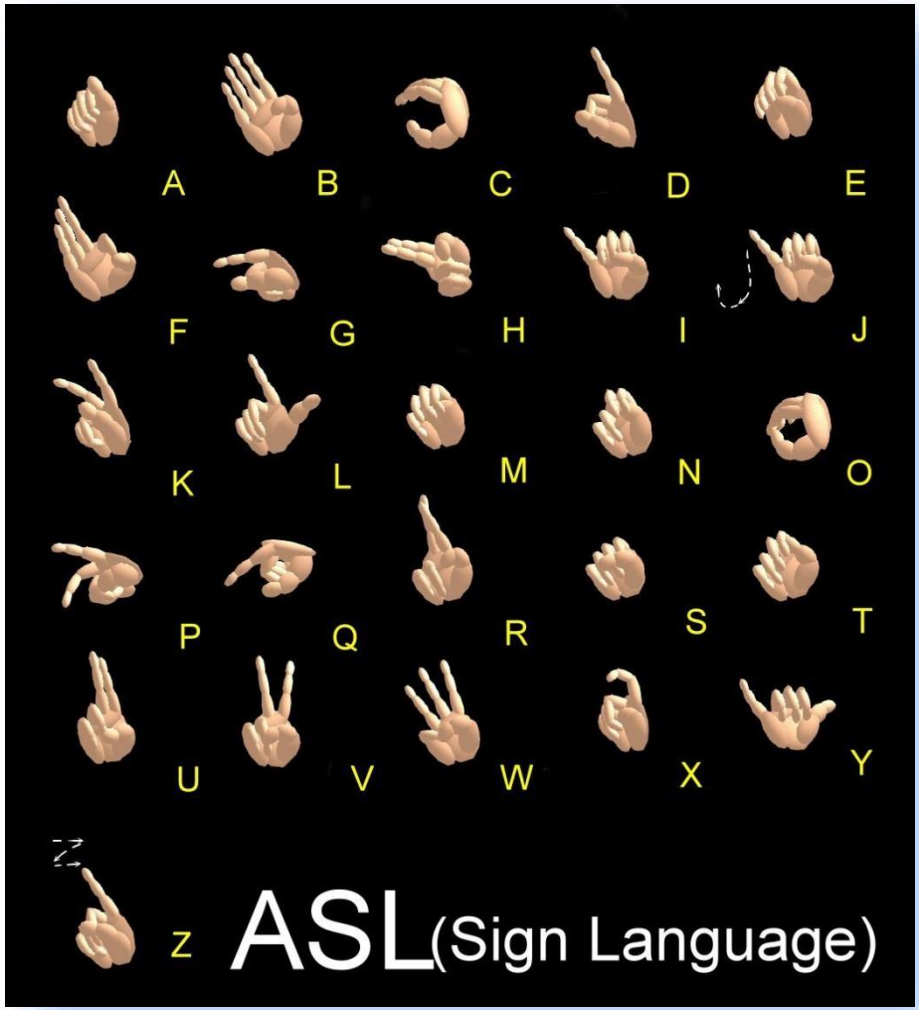
Tipuri de gesturi

- **Gesturi ale mâinii și brațelor**: recunoașterea ipostazelor mâinii, recunoașterea limbajului semnelor, transmiterea de comenzi, manevrare obiecte;

- Permite eliminarea contactului direct cu echipamentul manipulat.
- statice: ipostaze ale mâinii sau corpului;



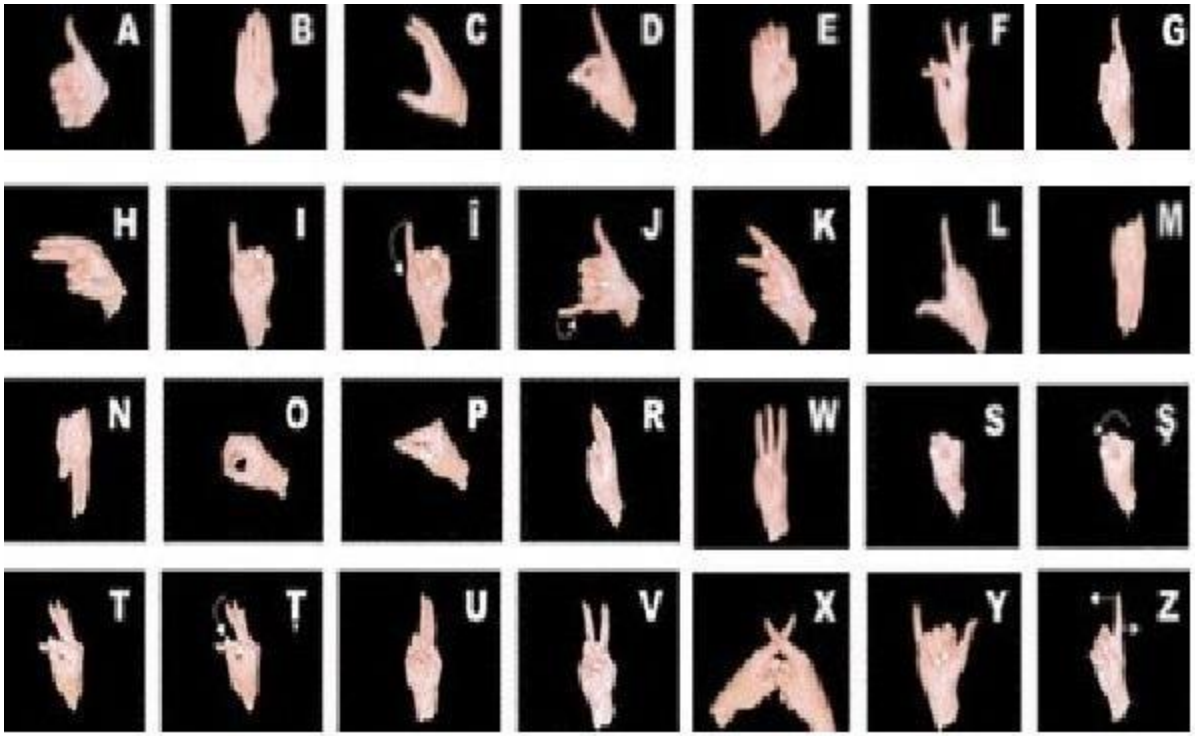
Tipuri de gesturi



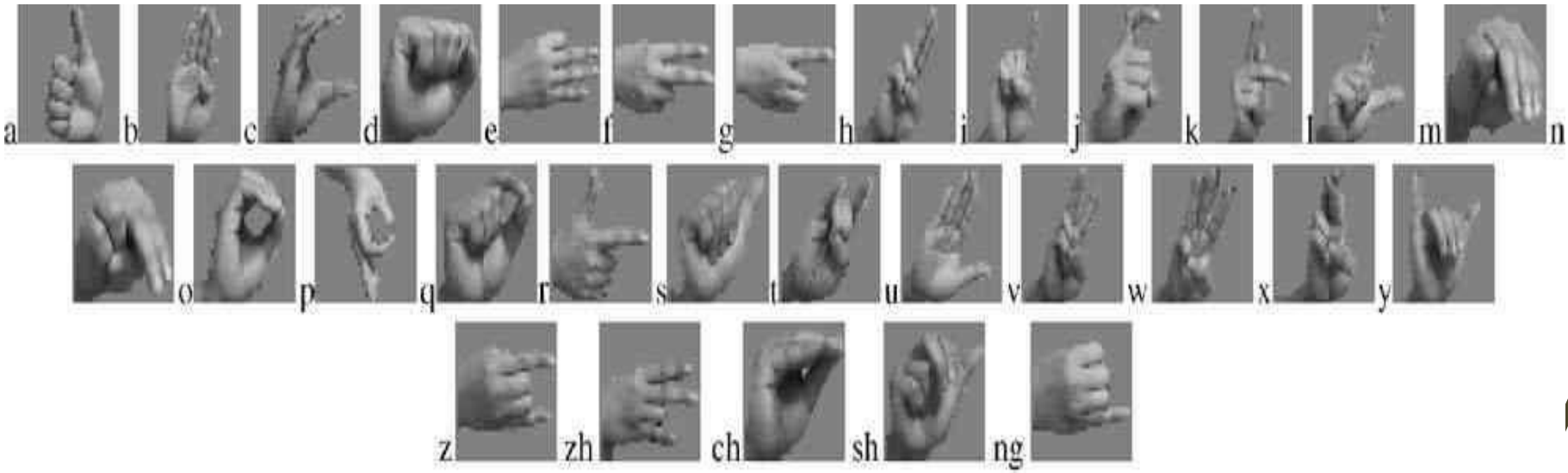
- American Sign Language (ASL)
  - Comunicare non-verbala
  - Contin si gesturi dinamice (J si Z)

Tipuri de gesturi

- Limbaj mimico-gestual romanesc



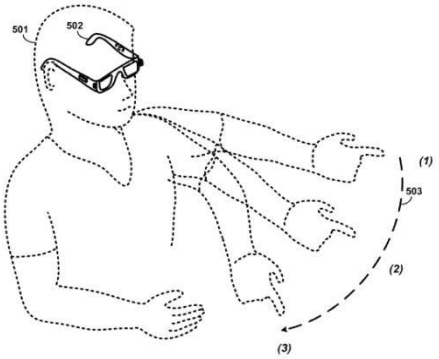
Tipuri de gesturi



- Limbaj mimico-gestual chinezesc

Tipuri de gesturi

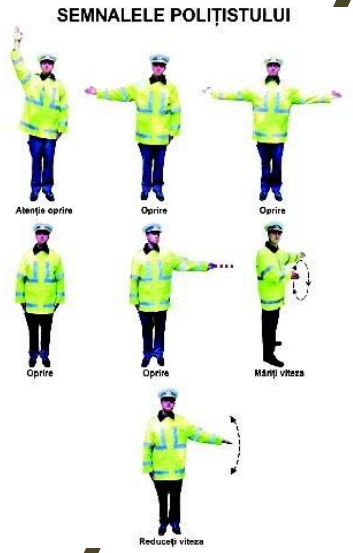
- **dinamice**: implică mișcare;



- **dinamice & statice**: implică mișcare în cadrul căreia sunt semnificative anumite ipostaze statice (ex: controlori trafic);







Tipuri de gesturi

- **Gesturi ale capului și feței**:
  - a da din cap
  - direcția privirii
  - ridicarea sprâncenelor,
  - deschiderea gurii pentru a vorbi,
  - clipire
  - expresii ale feței

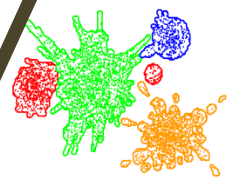


Tipuri de gesturi

- Gesturi ale corpului
  - Mod de interactiune a unor persoane
  - Urmarirea / corectitudinea miscarilor unui sportiv
  - Analiza cinematicii corpului



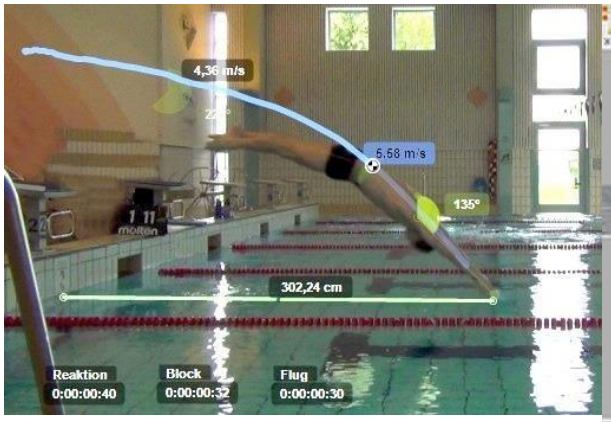


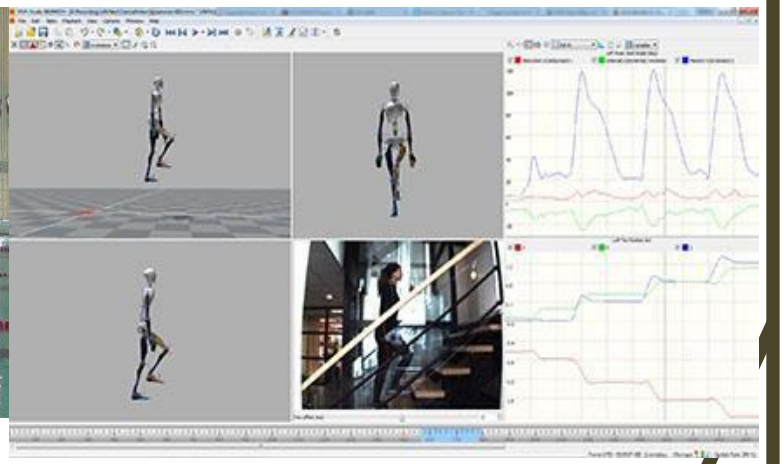


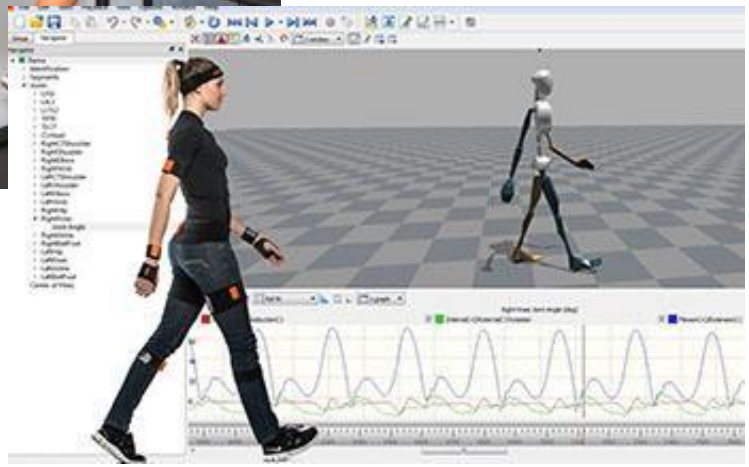
Tipuri de gesturi

- Imbunatatirea performantelor sportive







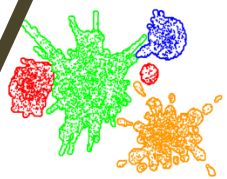




Tipuri de gesturi

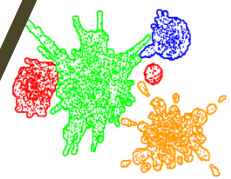
- Nu există o asociere unică între gest și informație (subiectivitate și repetitivitate limitată);

- Semnificația unui gest este determinată de:
  - informația spațială: unde are loc;
  - informația temporală: traiectoria în timp (“path”);
  - informația simbolică: semnificația;
  - informația afectivă: conținutul emoțional;



Provocari

- viteza - aplicațiile trebuie să fie adaptate să ruleze în timp real (minim 25 frame-uri pe secundă);
- acuratețe (robustete la schimbare de mediu / background);
    - fundalul de culoarea pielii poate interfera cu algoritmul de detecție
- robustețe la apariție de ocluziuni sau alte tipuri de obiecte din cadru;
- backlight;
- zgomot (camere de IR / RGB).



Provocari



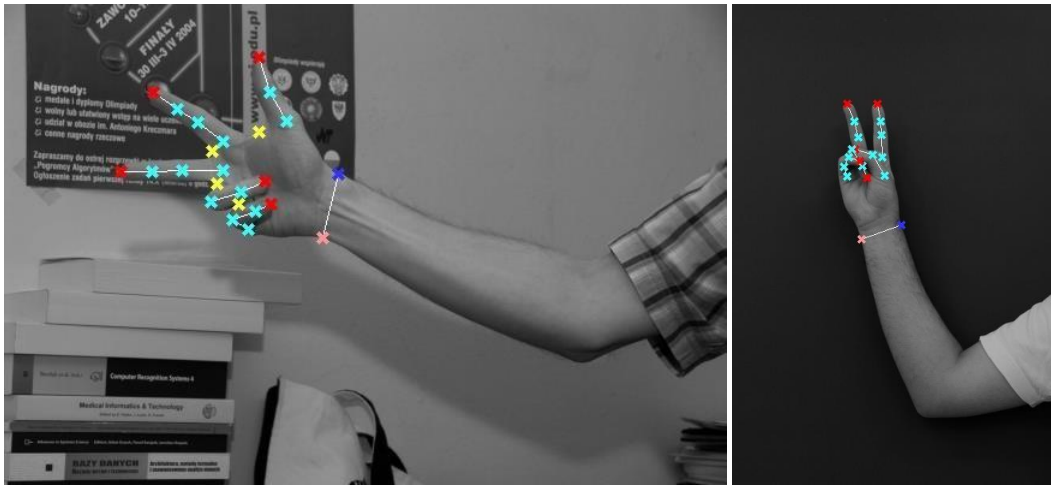
- **Sindromul „Gorilla arm”**

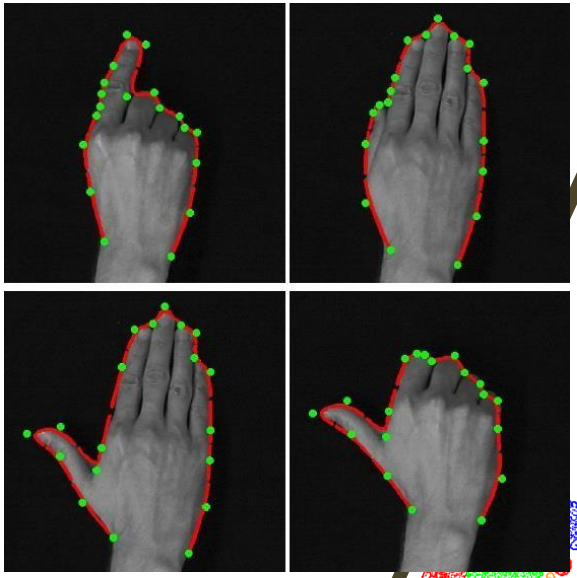
- După o utilizare îndelungată a mâinilor poate apare sindromul „Gorilla arm”.

- Oamenii nu pot „ține” mâinile în fața corpului pentru o perioada lungă de timp deoarece apar dureri, crampe și oferă impresia de mâine „supradimensionate”.

Clasificare

- **Fara contact**:
  - bazate pe prelucrarea imaginilor aferente preluate cu anumite dispozitive (camere video / stereo / adâncime etc.).





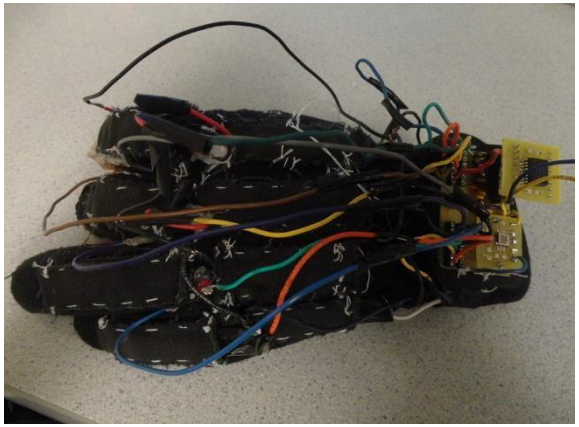
Clasificare

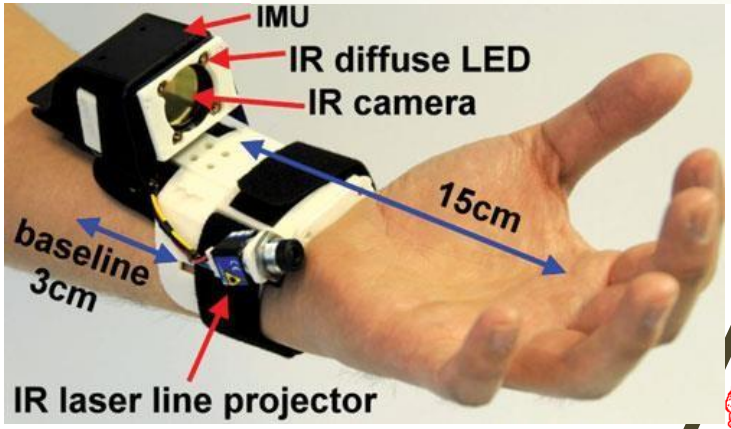
- **Cu contact**
  - bazate pe date preluate de către dispozitive harware specifice.







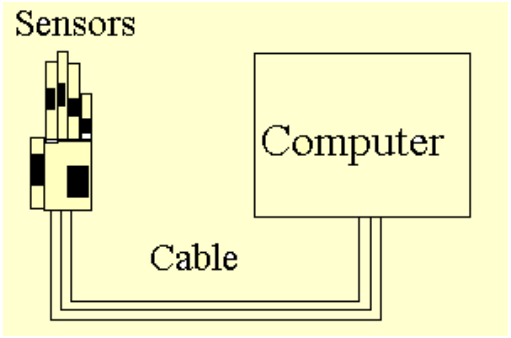




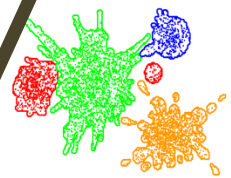
Exemple de dispozitive

- **Manusi**: Tehnologie de captare a mişcărilor (ex îndoirea degetelor).
- Conţin un detector de mişcare (magnetic sau inerţial) care captează poziţia globală a degetelor;
- Au un cost de producţie destul de ridicat;
- Sunt folosite în special pentru manipulare în medii virtuale și de “realitate augmentată”.









Exemple de dispozitive



- **Manusi „colorate”**
- Avantaje:
  - avem nevoie doar de un un webcam;
  - prețul mănușii este foarte mic (1USD).
- Dezavantaje:
  - probleme specifice atunci când se utilizează o cameră video ca și senzor (ocluziuni/zgomot).



Exemple de dispozitive

- Dispozitive atasate





WristQue



Fitbit



Cicret

Exemple de dispozitive

- Dispozitive atasate

Wii (Nintendo)



Myo





PS3 Move

Exemple de dispozitive

- **Camere de adâncime**

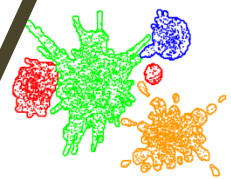


- Calculează distanţa dintre cameră şi obiectele din scena;
- Majoritatea calculează distanţa prin măsurarea distanţei pe care lumina o parcurge de la senzor la obiect.



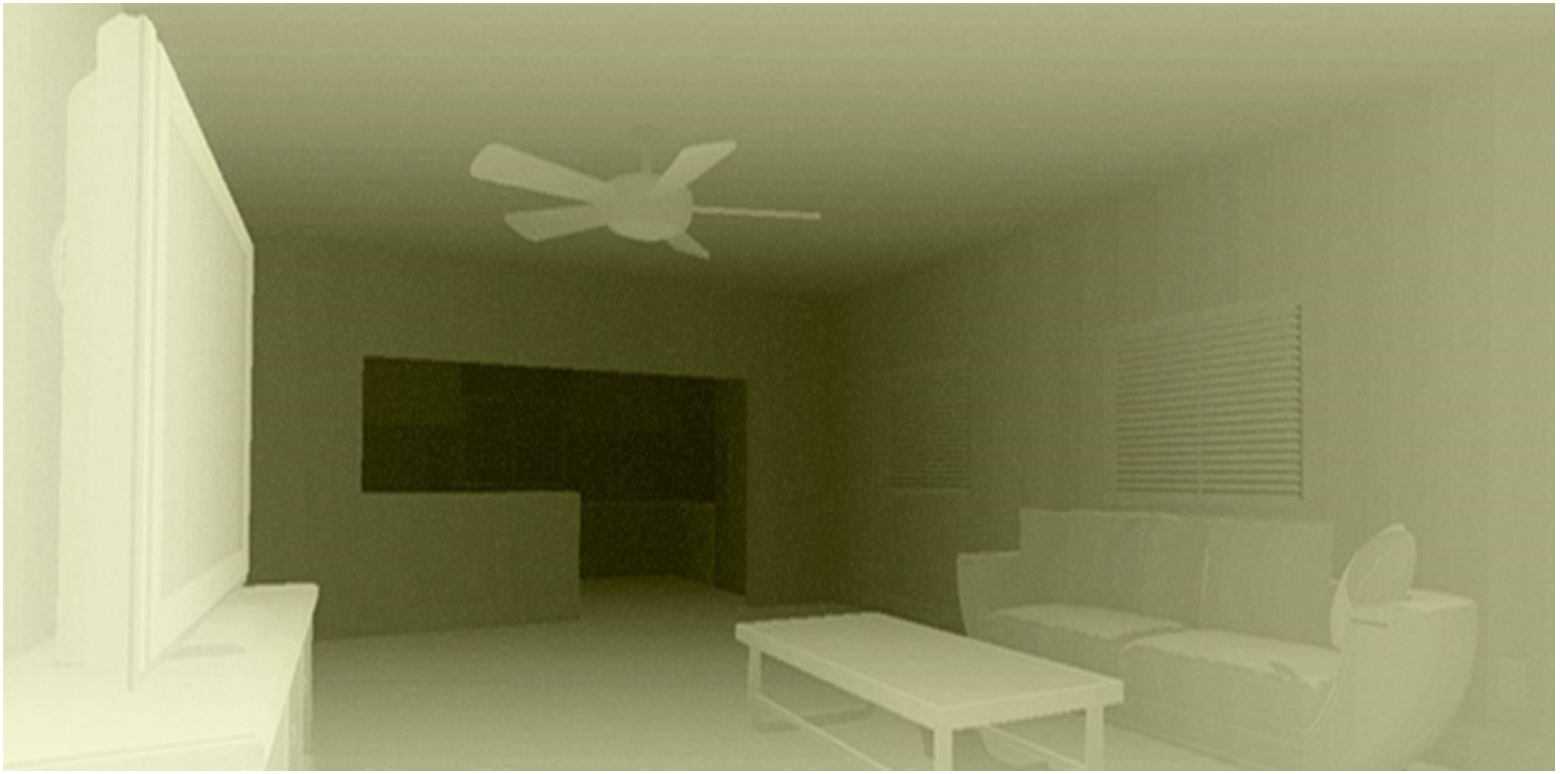




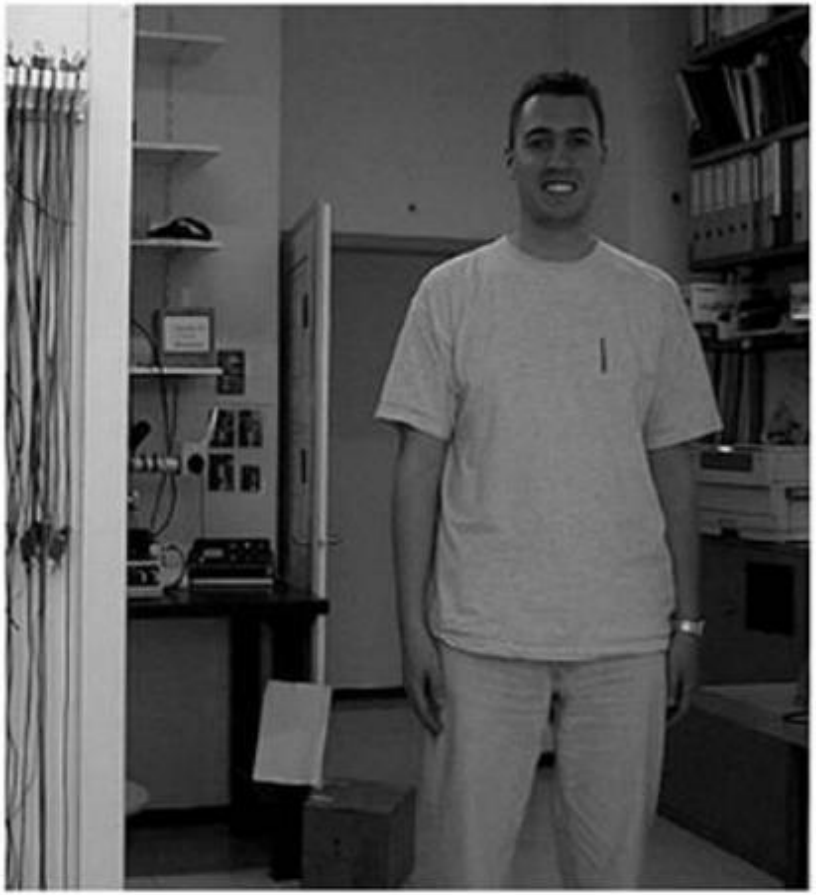


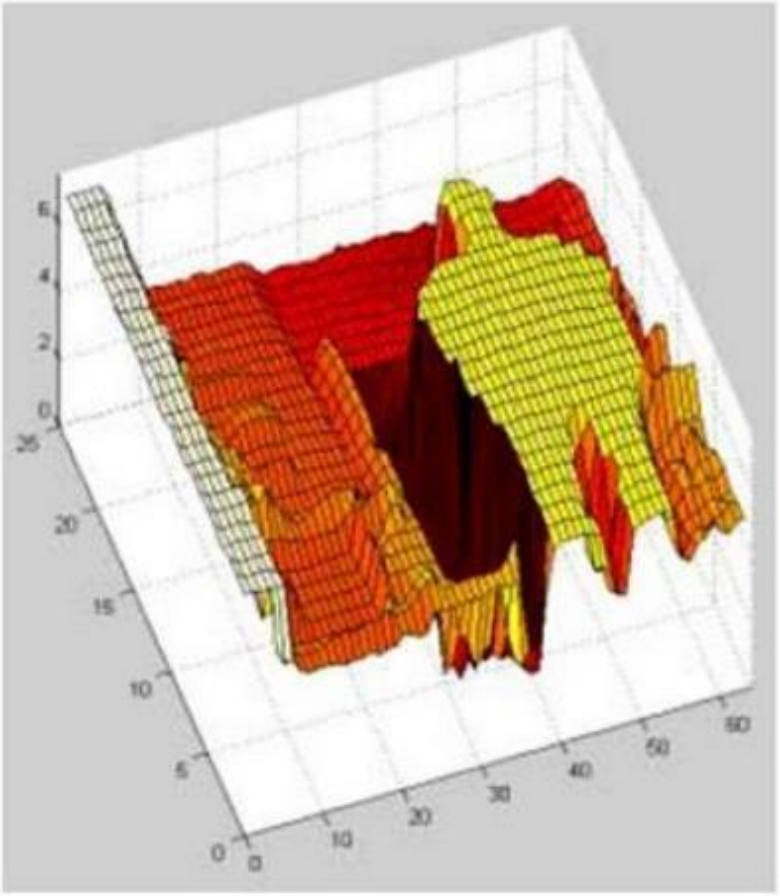
Camere de adancime

- Ce afișează o cameră de adâncime?



Camere de adancime



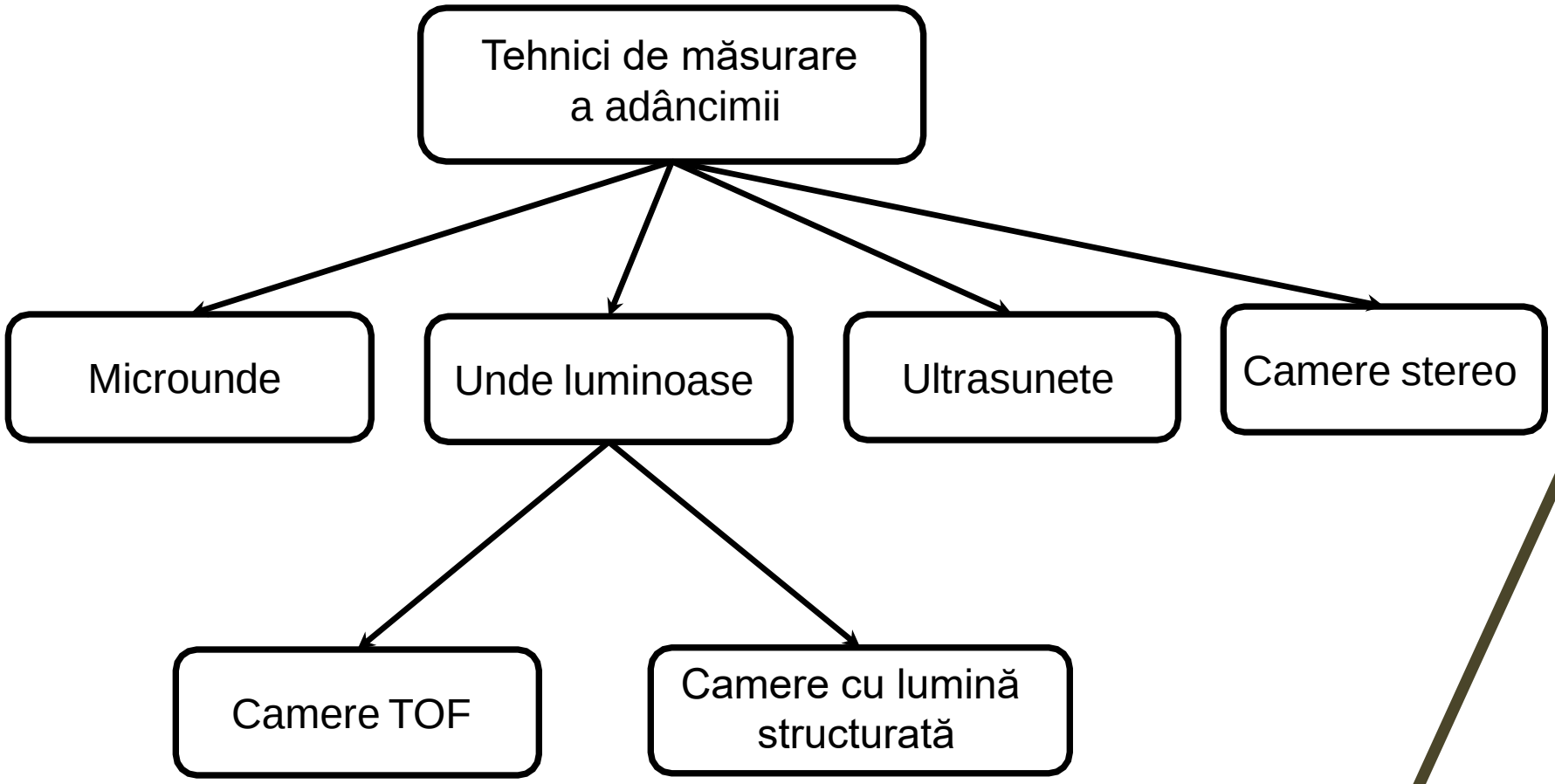


Camere de adancime





Camere de adancime



Camere stereo

- Reprezintă camere care au 2/3 lentile și care preiau mai multe imagini în același timp;
- Tehnologia a fost dezvoltată încă din anii 40;
- În acest mod se va simula vederea binoculară, capturând imagini tridimensionale;
- De obicei, distanța dintre lentile este aproximativ egală cu distanța dintre 2 ochi (distanța intra-oculară).



Camera Kodak



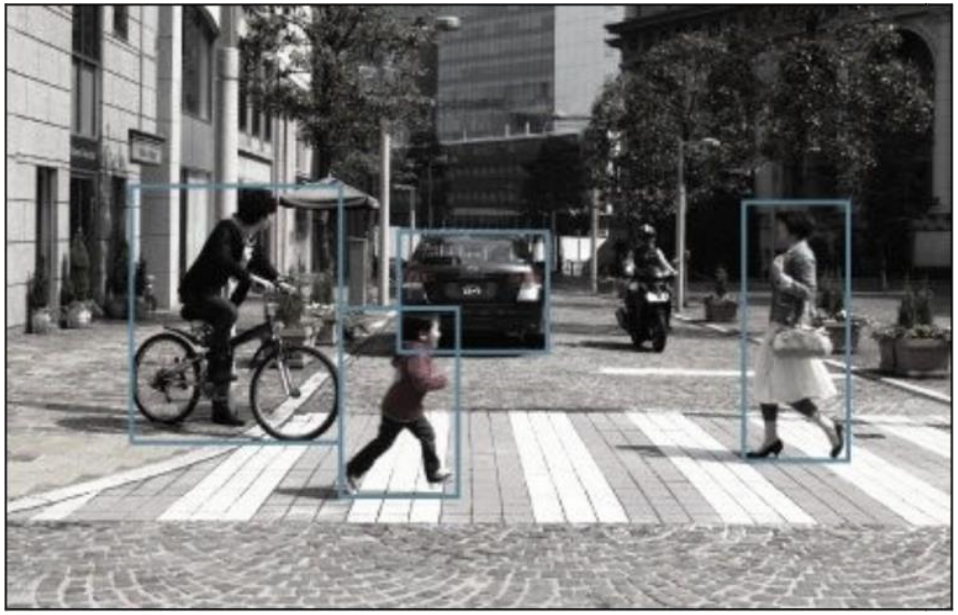
Camera Sputnik



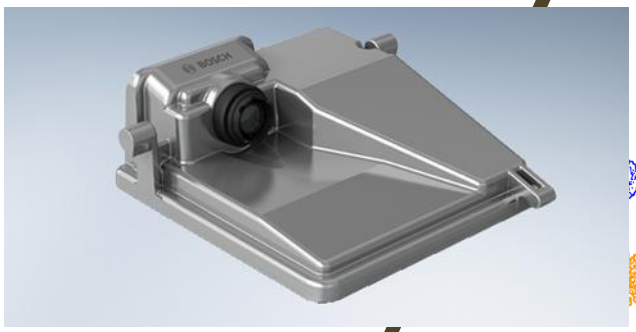
Camera Verascope

Camere stereo

- **Utilizare**
- Preluare de imagini și filme 3D (ex: Samsung NX300);
- montare camere stereo pe vehicule pentru a calcula direcția drumului și distanța față de pietoni (ex: Continental, Bosch, Subaru, Volvo, Nissan, Daimler).







Camere stereo

- **Utilizare**
  - Pentru navigarea roboților;

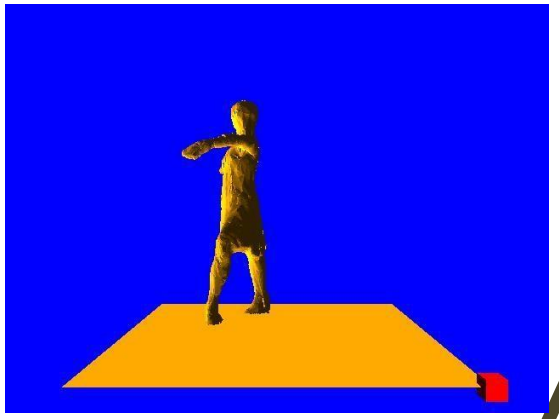


Ex: robotul Nomad caută meteoriți în Antartica

Camere stereo

- Reconstructie 3D;
  - Colecteaza flux video din 50 de camere





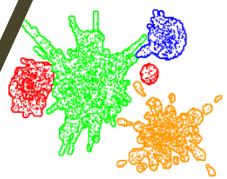
- http://www.cs.cmu.edu/afs/cs/project/VirtualizedR/www/VirtualizedR.html
  - [Takeo Kanade *et al.*, CMU]

Camere stereo



Camere stereo

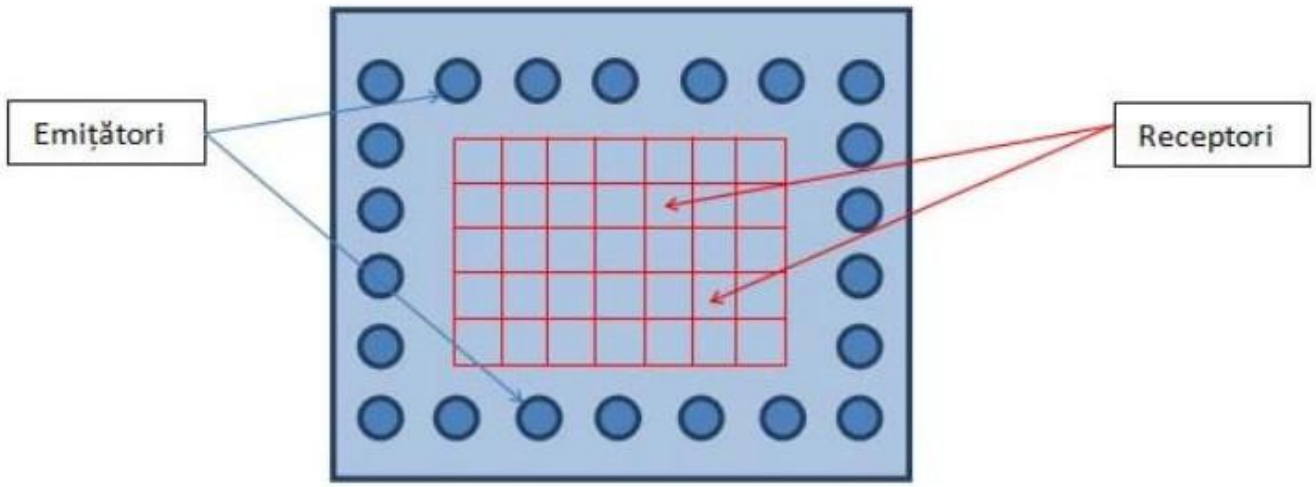
- **Dezavantaje**
  - Este nevoie de cel puțin două camere calibrate;
  - Costisitor din punct de vedere computațional;
  - Dependente de modul de iluminare a scenei.



Camere TOF

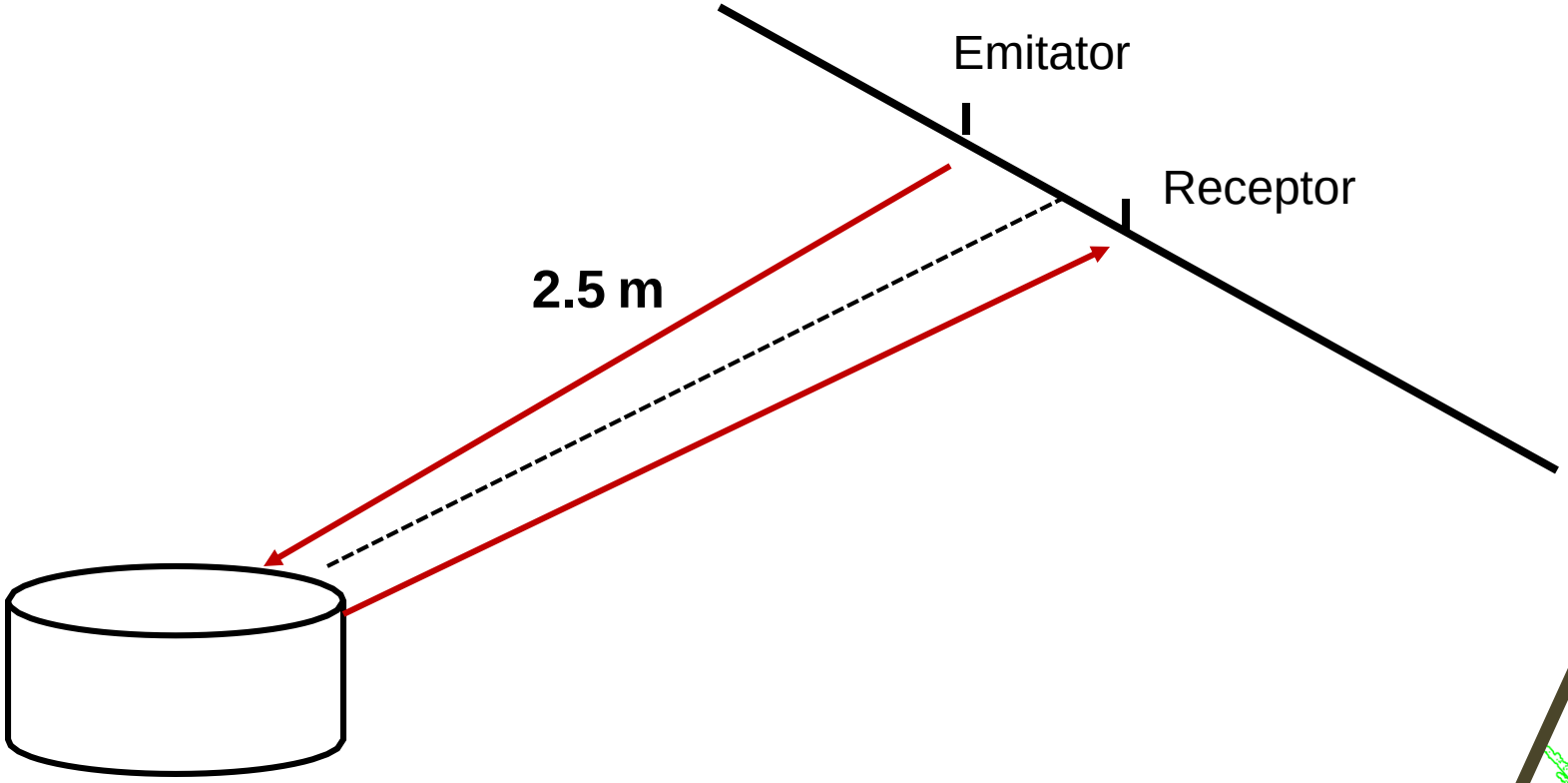
TOF = Time of Flight

- Camerele matriciale ToF estimează întreaga geometrie a scenei prin emiterea simultană a unui număr M x N fascicule luminoase care vor fi captate de un senzor;
- Fiecare punct din această matrice va indica în mod independent față de celălalte puncte distanța de la punctul corespunzător al matricei la un punct corespunzător din scenă.



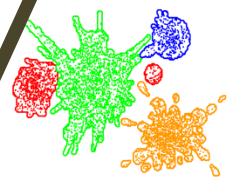
Camere TOF

- Se emit fascicule de lumină pe perioade foarte scurte



Camere TOF

- **Simplitate în implementare**
  - ✓ sistem compact (față de sistemele stereo sau cu triangulizare);
  - ✓ nu este nevoie de module mecanice (în contrast cu sisteme cu scanare laser);
  - ✓ algoritmul de calcul al distanței este eficient (putere de procesare mică);
  - ✓ este nevoie de o singură cameră calibrată.
- **Viteză**
  - ✓ capabile să măsoare distanțele din încăpere într-o singură iterație;
  - ✓ unele dispozitive pot atinge și 160 cadre pe secundă, ceea ce le face ideale pentru aplicații în timp real.
- **Preț redus**



Camere TOF

- **Dezavantaje**

➢ **Background backlight**

➢ **Interferențe**

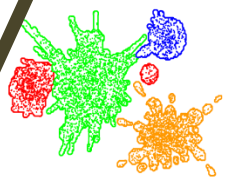
✓ mai multe dispozitive TOF similare pot interfera;

✓ se pot elimina prin anumite tehnici:

- multiplexare temporală (doar un dispozitiv emite la un moment dat);
- frecvențe de modulație diferite.

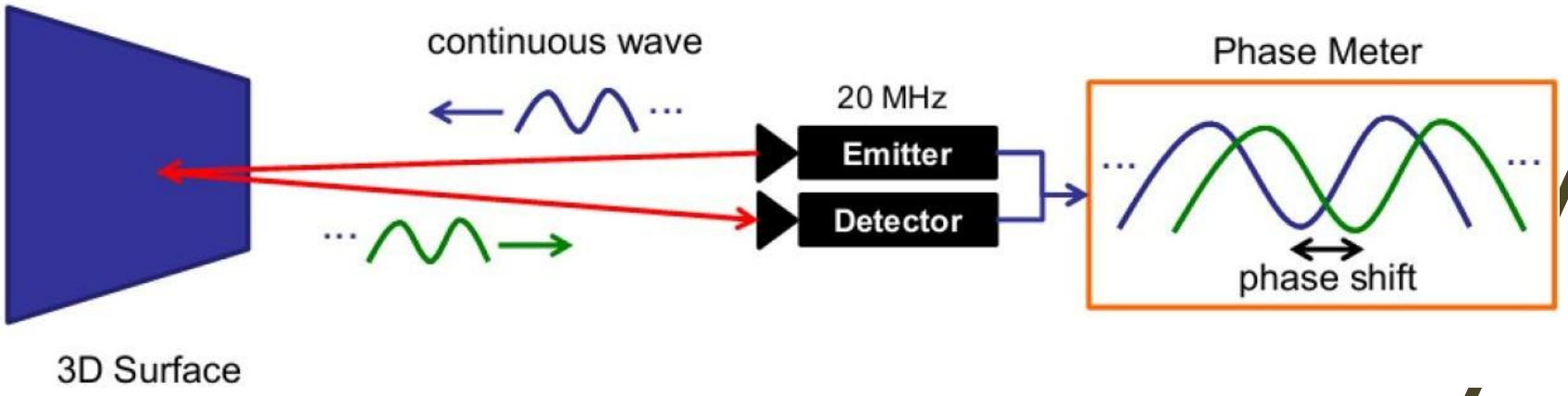
➢ **Reflexii**

✓ dispozitivele TOF iluminează întreaga scenă, iar datorită reflexiilor lumina poate „atinge” obiectele urmând mai multe căi posibile, iar distanța estimată poate fi mai mare.



Camere TOF

- ✓ Sunt emise unde luminoase continue sinusoidale (în loc de pulsuri);
- ✓ Unda luminoasă primită va avea o diferență de fază care va fi proporțională cu distanța senzor – obiect.



Kinect

- Kinect-ul reprezintă un dispozitiv de recunoaștere de mișcare disponibil pe:
  - Xbox 360 / One
  - PC
- Ajută utilizatorul să interacționeze / controleze cu jocurile și aplicațiile:
  - fără a avea nevoie să atingă tastatura / mouse-ul / joystick
  - reprezintă o interfață naturală om – mașină;
  - se pot utiliza gesturile și comenzi vocale.





Kinect

Emitator IR

Senzor de adâncime (receptor IR)

-măsoară distanța față de obiectele din cadru

-captează informația de mișcare

Camera RGB

se poate antrena pentru detecție de fețe



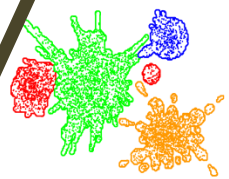
Microfon Multi-array

- permite recunoașterea si urmărirea sursei audio / recunoaștere de voce

Kinect

- **Ce își propune să rezolve un Kinect?**

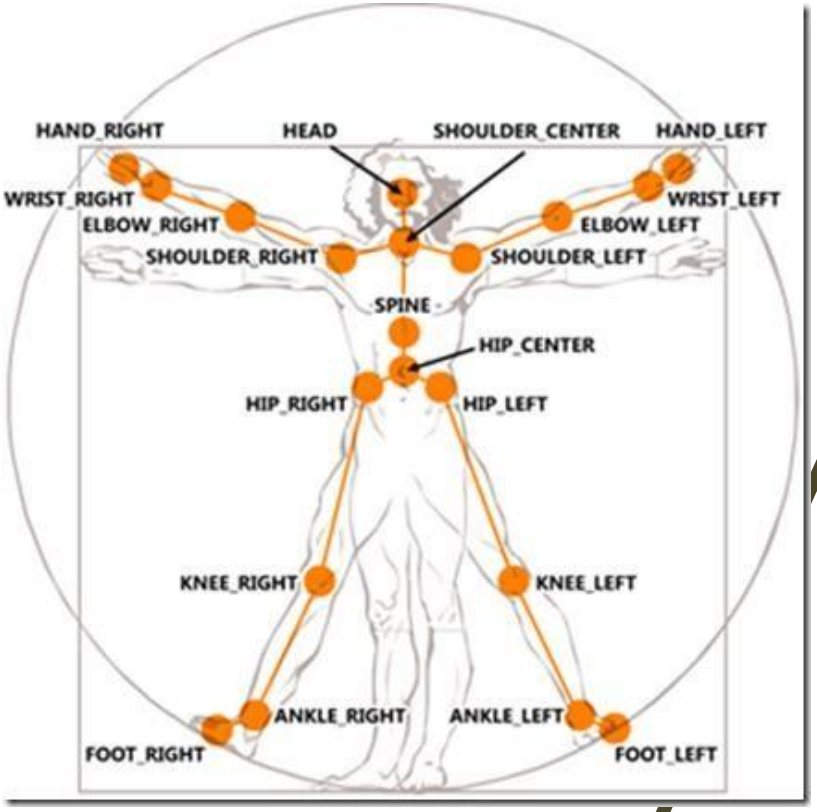
- detecție persoane în scenă ignorând informația de fundal;
- detecție părți componente ale mâinii;
- detecție și urmărire de gesturi;
- un prim pas către recunoaștere de fețe;
- recunoaștere de comenzi vocale.

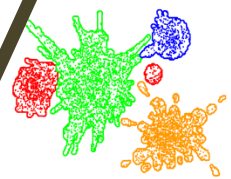


Kinect

- **Recunoaștere skeleton**

- Se pot urmări 1-2 / 5-6 jucători;
- Se calculează coordonatele xyz față de cameră;
- Stări:
  - tracked, not tracked, inferred.



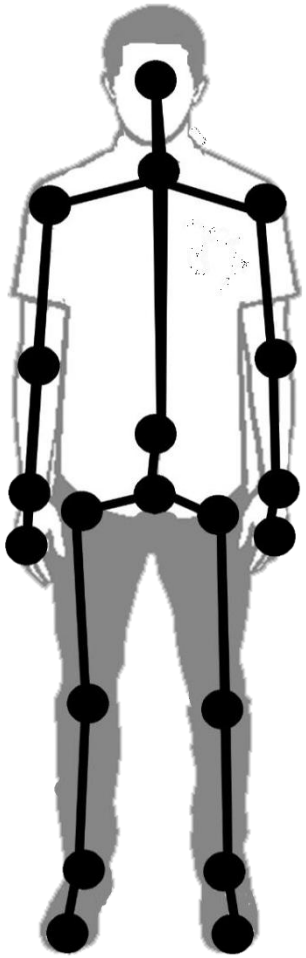


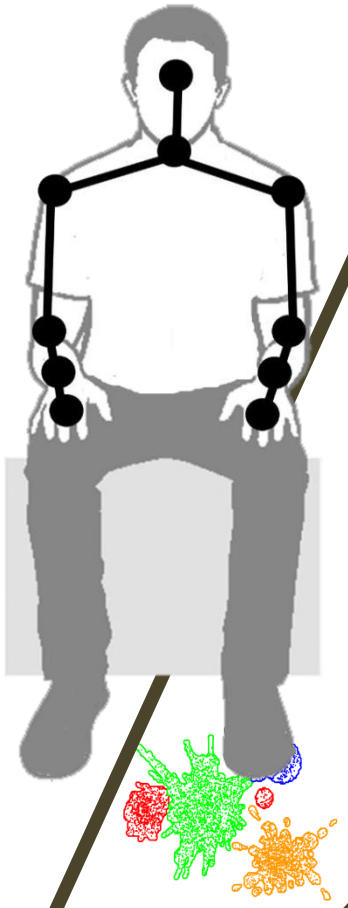
Kinect

- Recunoaștere skeleton

- 20 / 25 puncte de urmărire (ridicat);
- 10 puncte de urmărire (așezat).

- Părțile componente sunt deduse utilizând algoritmul „random forests” care a fost antrenat utilizând peste 1 milion de exemple de antrenare.

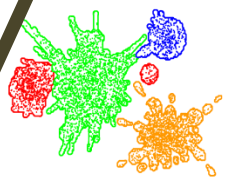




Recunoasterea de gesturi prin analiza de imagini si video

Selectia punctelor importante

Interfata vizuala om-masina



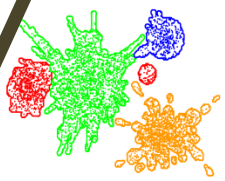
Introducere

Urmarire liniilor de demarcatie - automotives



Apostoloff, N. E. and A. Zelinsky. *Vision in and out of vehicles: integrated driver and road scene monitoring*. International Journal of Robotics Research, Volume 23, Number 4, pp. 513-538, 2004

http://www.robots.ox.ac.uk/~nema/research/lane%20tracking/lane_tracking.html

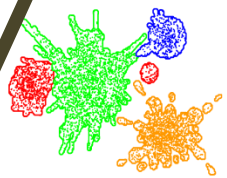


Introducere

Urmarire mainii – recunoasterea gesturilor



Blake, A., Isard, and M., “Active contours”, Springer, 1998
http://www.robots.ox.ac.uk/~ab/dynamics.html

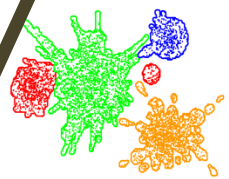


Introducere

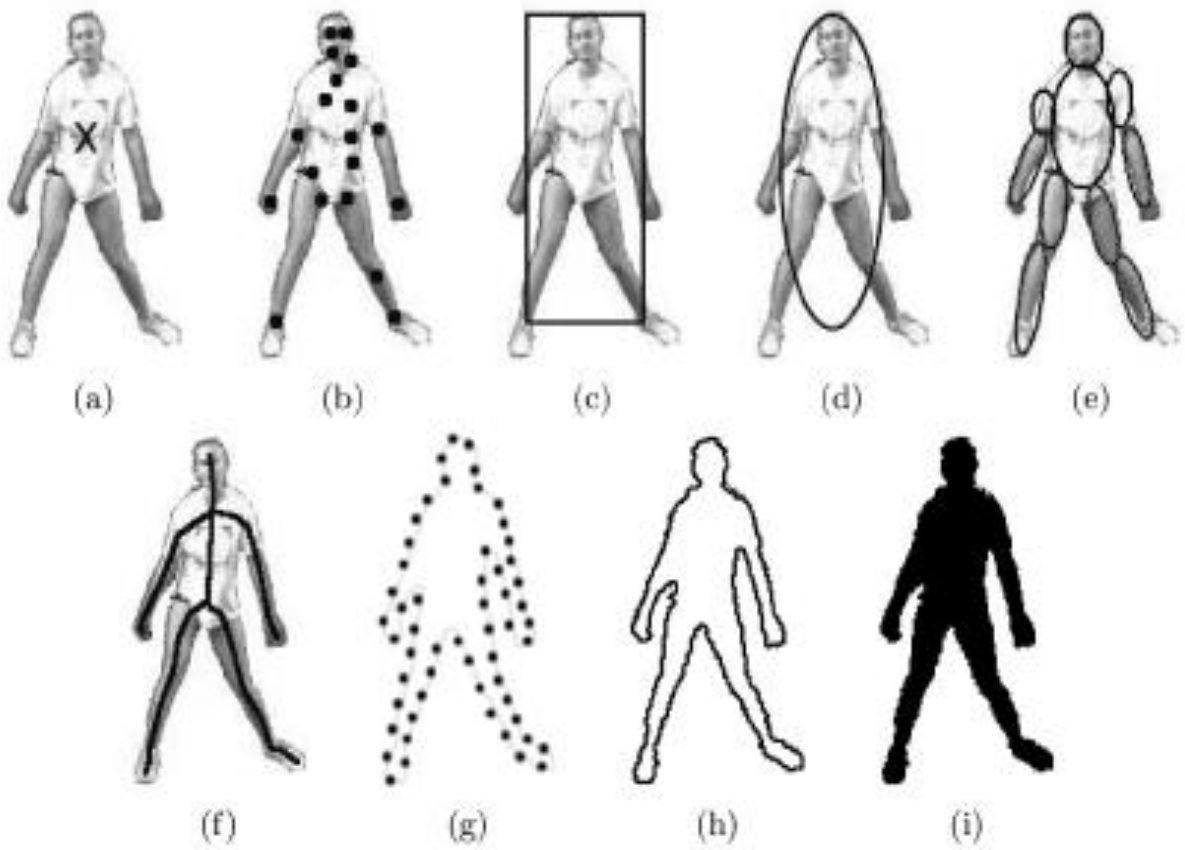
Miscare rapida si haotica – anticiparea miscarii



Blake, A., Isard, and M., “Active contours”, Springer, 1998
http://www.robots.ox.ac.uk/~ab/dynamics.html



Reprezentarea obiectului



(a) Punct

(b) Set de puncte

(c) Dreptunghi de incadrare

(d) Elipsa

(e) Set de elipse

(f) Skeleton

(g) Puncte de control pe contur

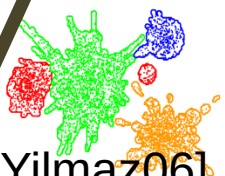
(h) Contur complet

(i) Silueta

Reprezentarea prin contururi sau siluete - obiecte complexe, nerigide

Reprezentarea prin forme geometrice - obiecte rigide (forma fixa)

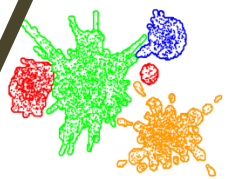
Reprezentarea prin puncte - obiecte mici



[Yilmaz06]

Introducere

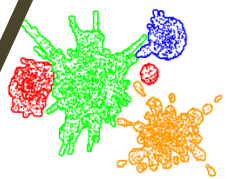
- Pasi:
  - Gasirea de puncte de interes
  - Imperecherea punctelor gasite
  - Urmarirea miscarii
  - Decizie pe baza modului de miscare



Gasirea punctelor de interes

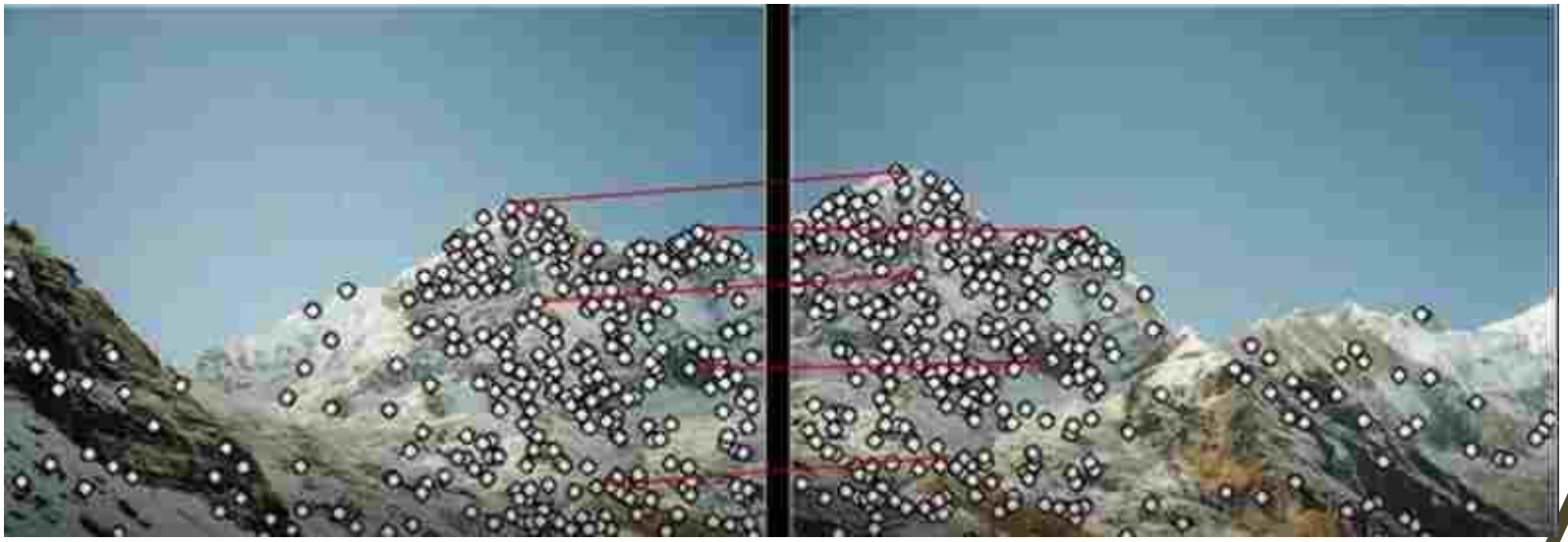
- Punerea problemei:
  - 2 imagini ce contin obiectul in pozitii diferite
  - (problema similara) doua imagini care sunt legate printr-o transformare geometrica (la vedere stereo, la panorame)

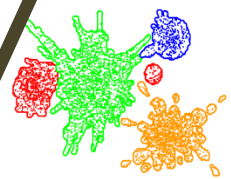
- Cum gasesc functia de transformare intre cele doua imagini?



Gasirea punctelor de interes

- Scop: sa gasim, intr-o imagine, puncte care:
  - Pot fi gasite si in a doua imagine
  - Pot fi localizate cu exactitate
  - Pot fi imperecheate unic cu punctele din a doua imagine





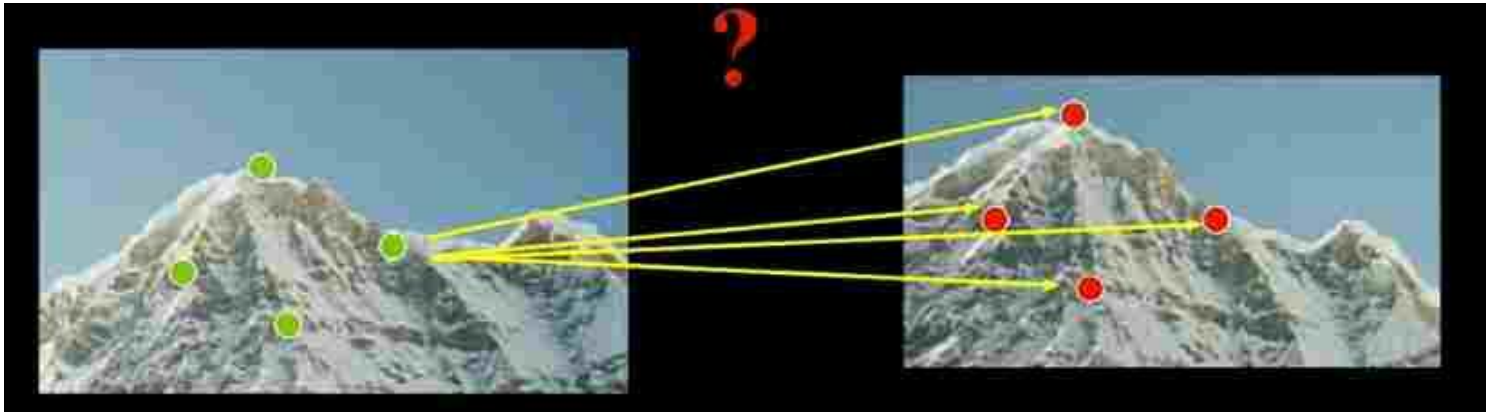
Gasirea punctelor de interes

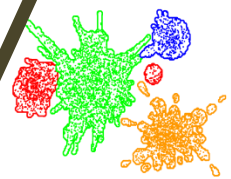
- Localizeaza trebuie sa fie repetabila -> sa localizeze aceleasi puncte ca fiind de interes





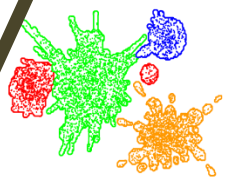
- Descrierea punctului trebuie sa fie (ideal) ‘unica’





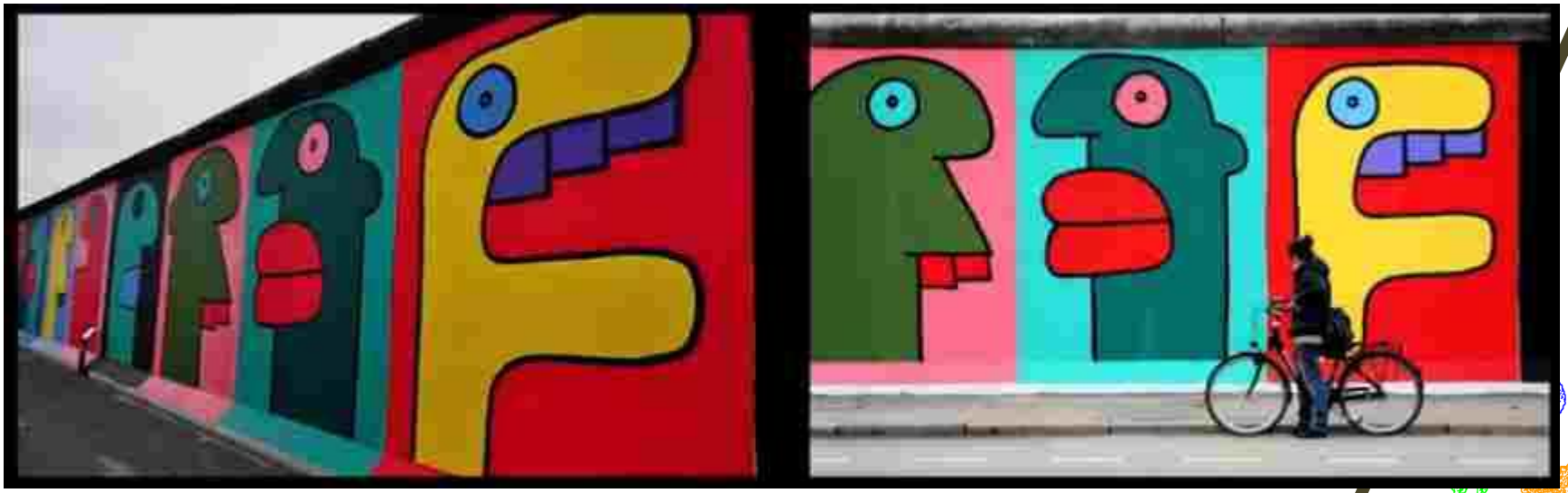
Gasirea punctelor de interes

- Repetabilitate / precizie
- Unicitatea descrierii
- Eficienta / numar mic de puncte
- Descrierea este facuta local - pe o vecinatate cat mai mica (pentru a evita ocluziile)



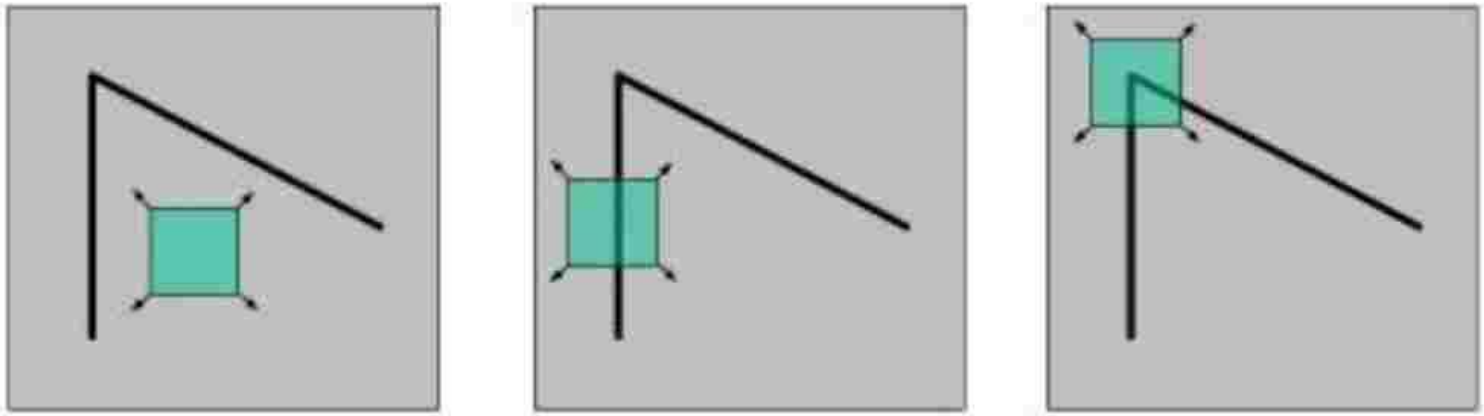
Gasirea punctelor de interes

- Punctele de interes pot fi:
  - In zone uniforme?
  - In zone de contur?
  - In zone de colt?



Detectie de colturi: metoda Harris

- Idee: in zonele de colt, gradientii pe orizontala si verticala se modifica mult



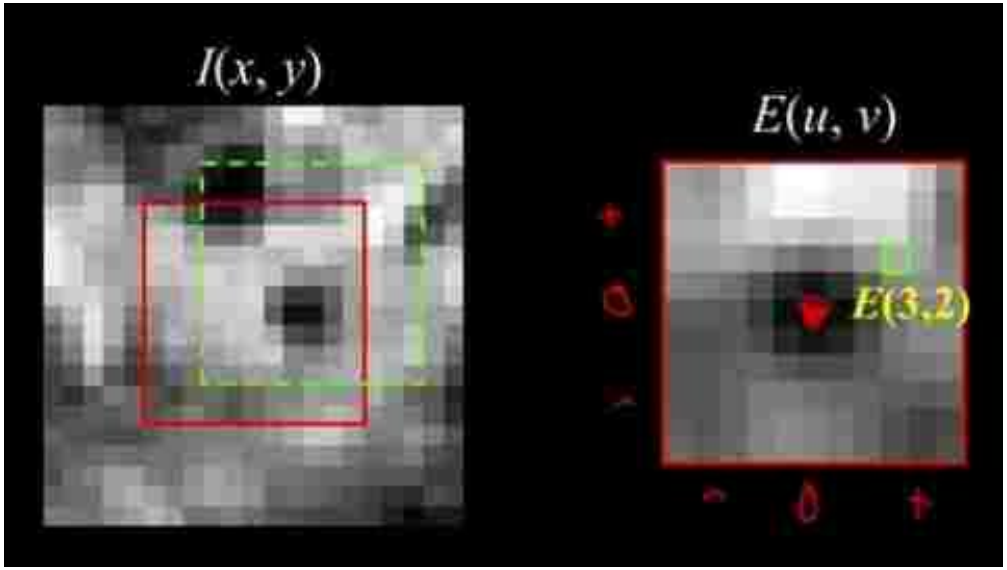
Chris Harris and Mike Stephens "A Combined Corner and Edge Detector“, *Alvey Vision Conference, 1988*

Metoda Harris

- Calculam:

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

Unde w(x,y) pot fi ponderi egale (medie aritmetica), gaussiene, etc.



Vrem sa maximizam functie E(u,v).

Metoda Harris

- Matematica pe scurt:
  - E(u,v) se dezvolta in serie Taylor in jurul lui (0,0) – aproximare de ordin 2
  - Se calculeaza derivatele partiale
  - Se ajunge la:

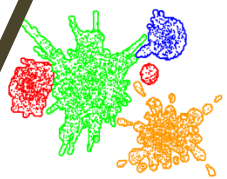
$$E(u,v) \approx [u \quad v]\left(\sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}\right)\begin{bmatrix} u \\ v \end{bmatrix}$$

Unde I_x si I_y sunt derivatele pe x si y ale imaginii

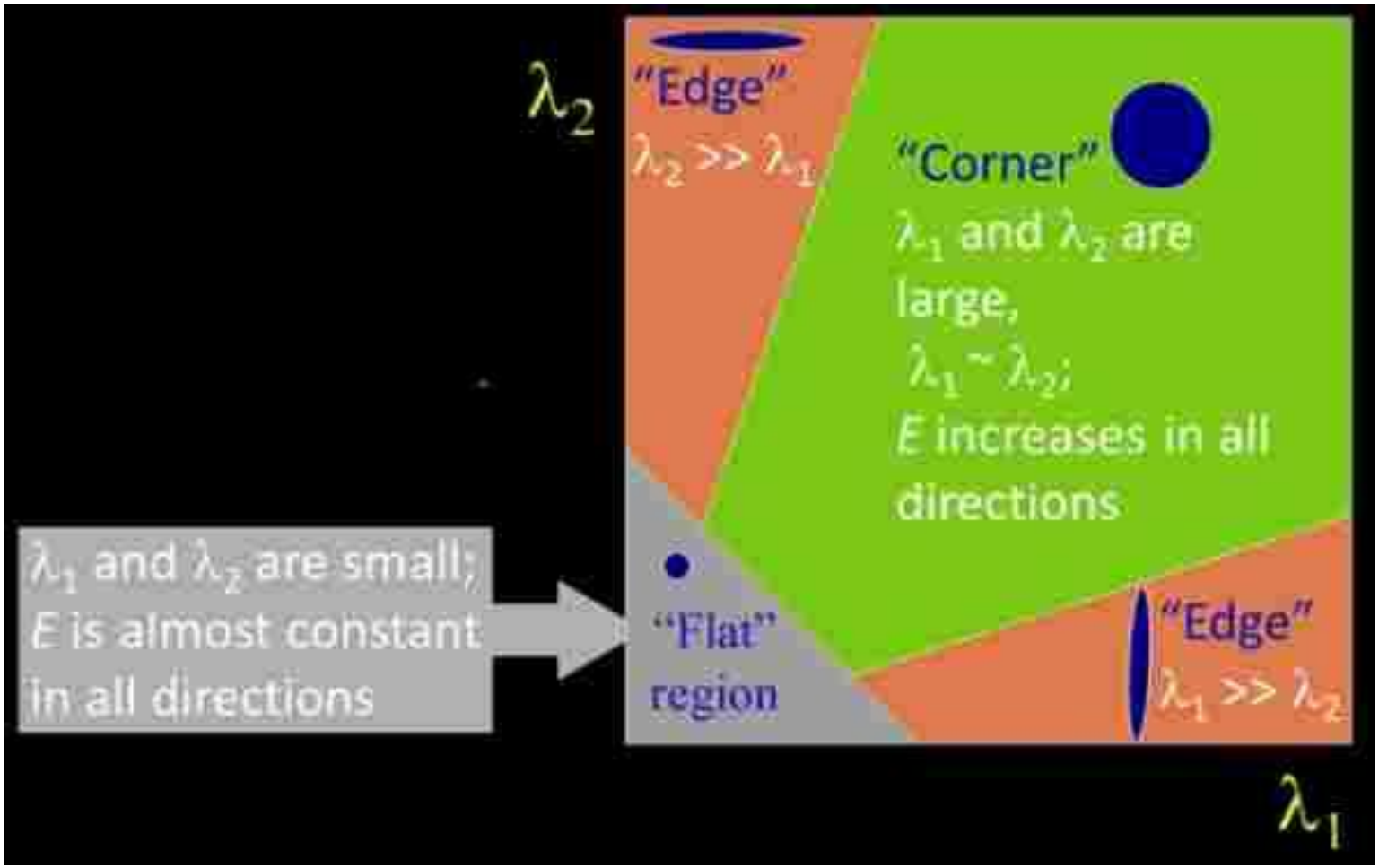
  - Se noteaza:

$$M = \sum w(x,y)\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \longrightarrow E(u,v) \approx [u \quad v]\, M \begin{bmatrix} u \\ v \end{bmatrix}$$

  - Se calculeaza valorile si vectorii proprii a lui M
  - Valorile proprii ne dau informatii despre E(u,v)



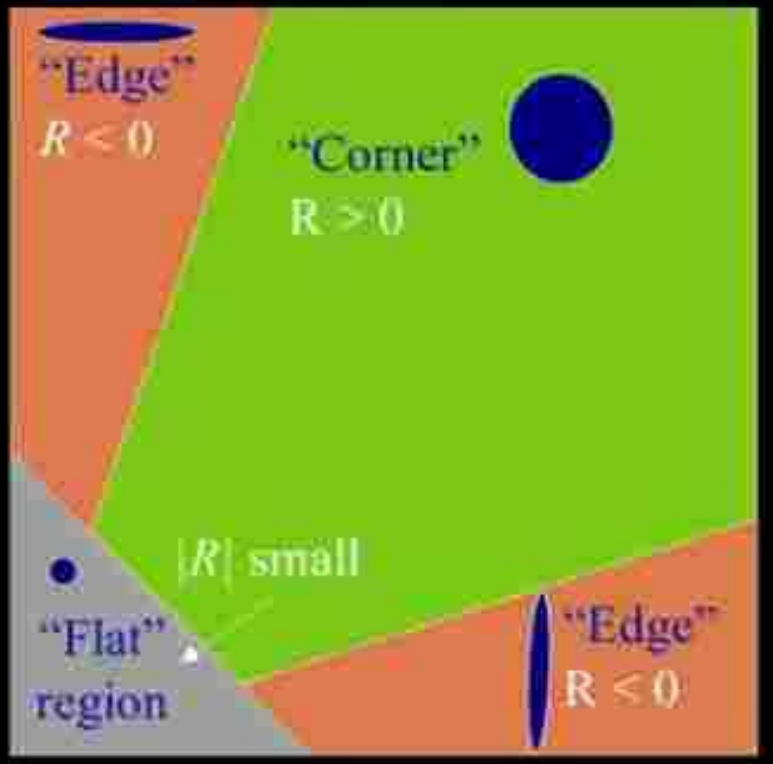
Metoda Harris



Metoda Harris

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

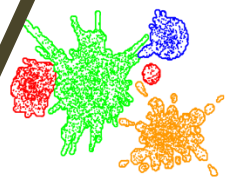
α: constant (0.04 to 0.06)



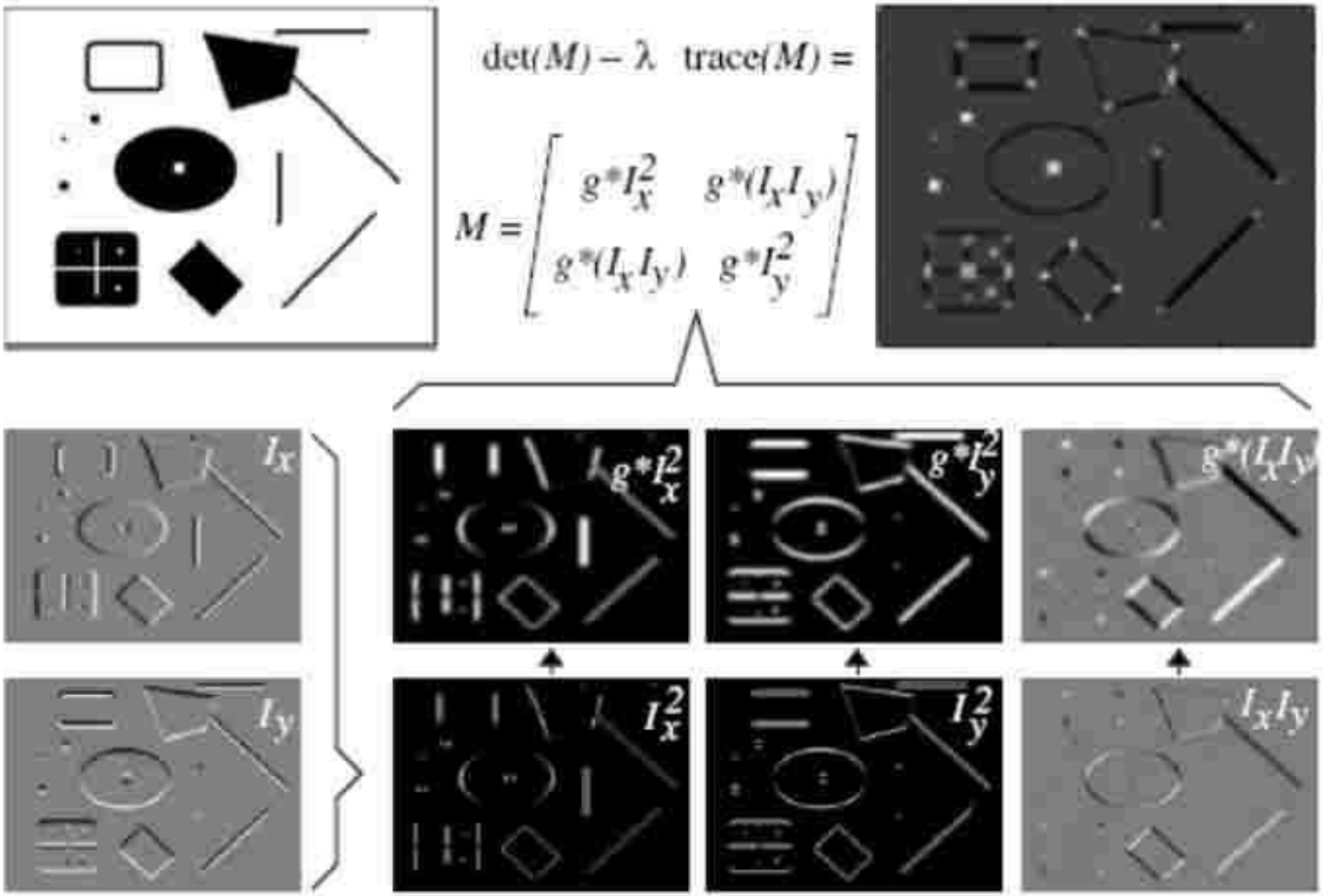
Metoda Harris

Algoritm:

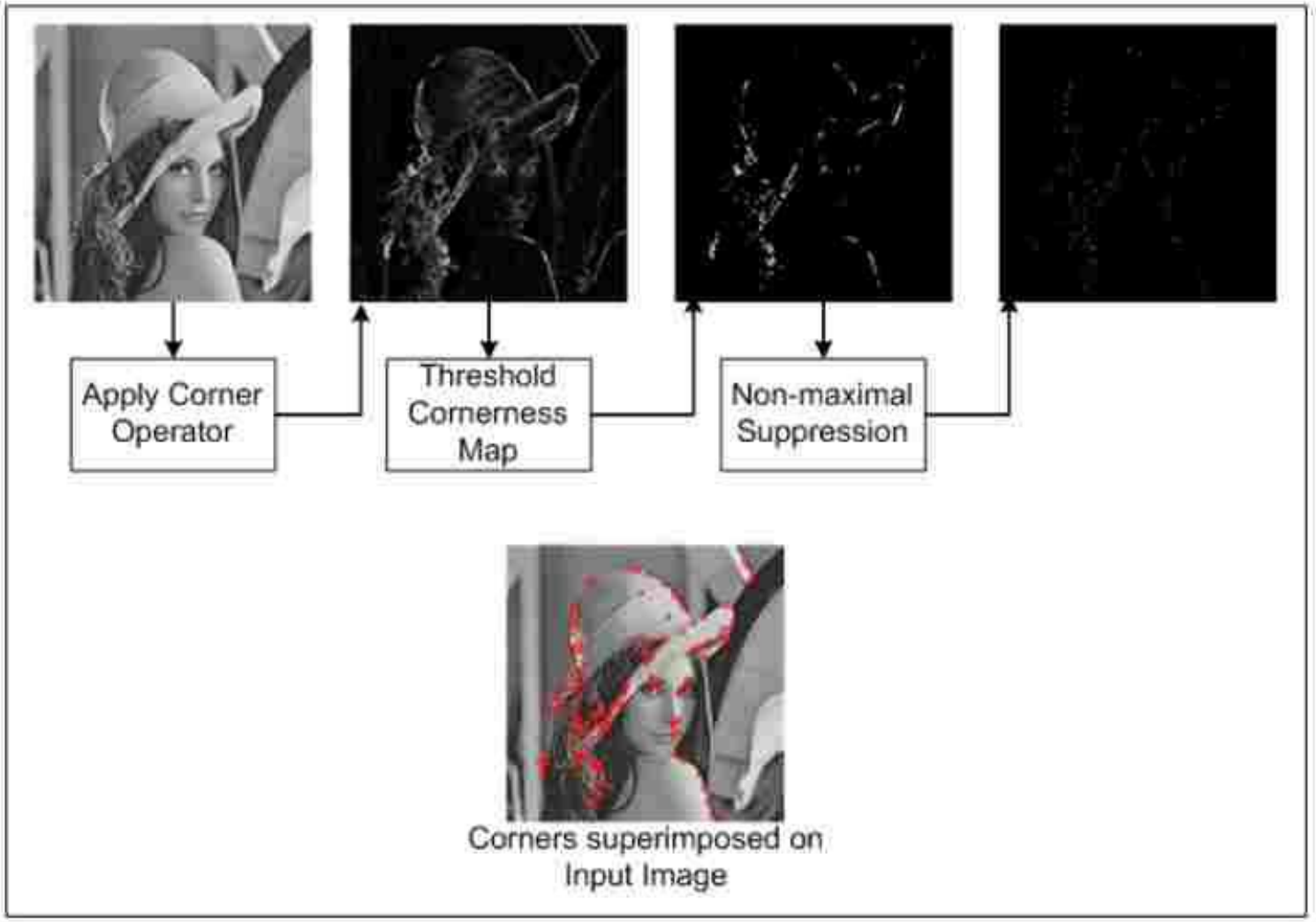
- Calculeza I_x si I_y (derivatele pe x si y ale imaginii)
  - Cu masca Sobel (pt. a include si blurare gaussiana)
- Calculeaza M
- Calculeaza R
- Gaseste pixelii in care R e mai mare decat un prag si este maxim local in regiune (local non-maximum suppression)



Metoda Harris



Metoda Harris



Metoda Harris

Metode similare:

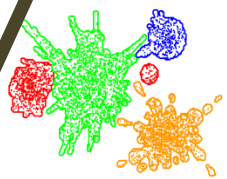
- Metoda Shi-Tomasi

  Cornerness = min (λ_1, λ_2)

- Metoda Brown-Szeliski- Winder

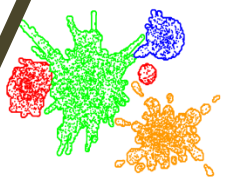
  Cornerness = det(M) / trace(M)

J. Shi and C. Tomasi "Good Features to Track*", CVPR,* pp. 593–600*,* 1994
M. Brown, R. Szeliski and S. Winder, "Multi-image matching using multi-scale oriented patches," *CVPR',* pp. 510-517, 2005.



Metoda Harris

- Invarinta la rotatie?
  - Da (valorile proprii raman la fel)
- Invarianta la schimbari de luminozitate?
  - Da pentru schimbari aditive (M depinde numai de derivate, care raman neschimbate).
  - Partial pentru schimbari multiplicative (derivatele si implicit R se vor modifica proportional, punctele cu valori R aproape de prag pot disparea, sau pot aparea puncte noi
- Invarianta la scalare?
  - **Nu**, chiar si cu factori de scalare mici (2,3…).

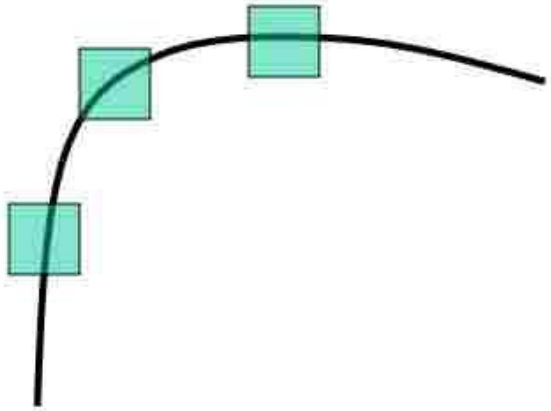


Metoda Harris

- Scaling

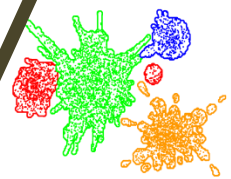


Corner



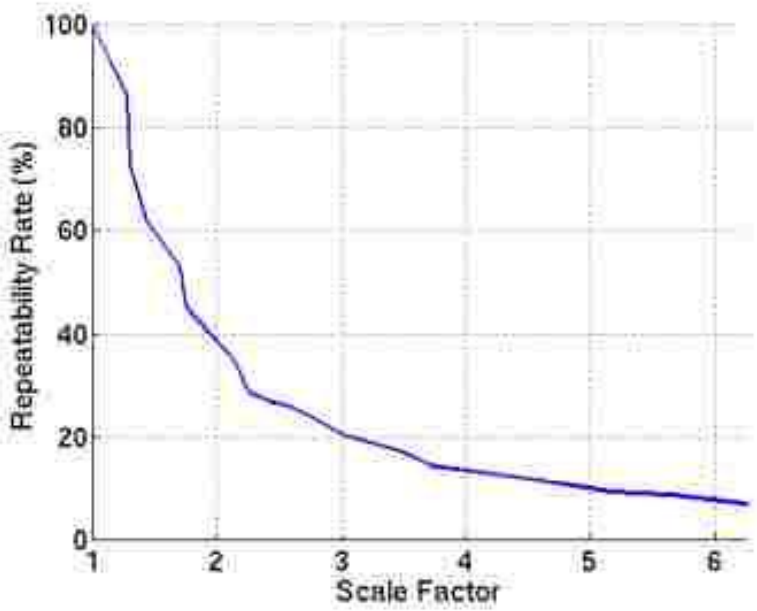
All points will be classified as edges

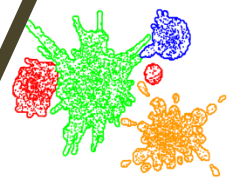
Not invariant to scaling (and affine transforms)



Metoda Harris

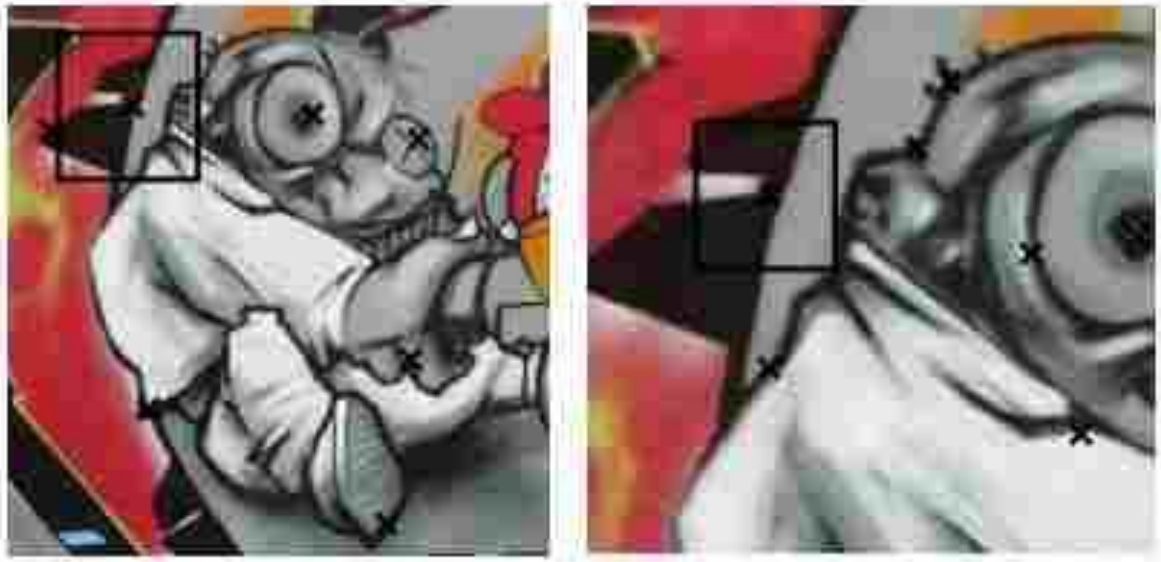
- Sensibila la:
  - Rotatii
  - Schimbari majore de perspectiva
  - Schimbari majore de contrast





Metoda Harris multi-scala

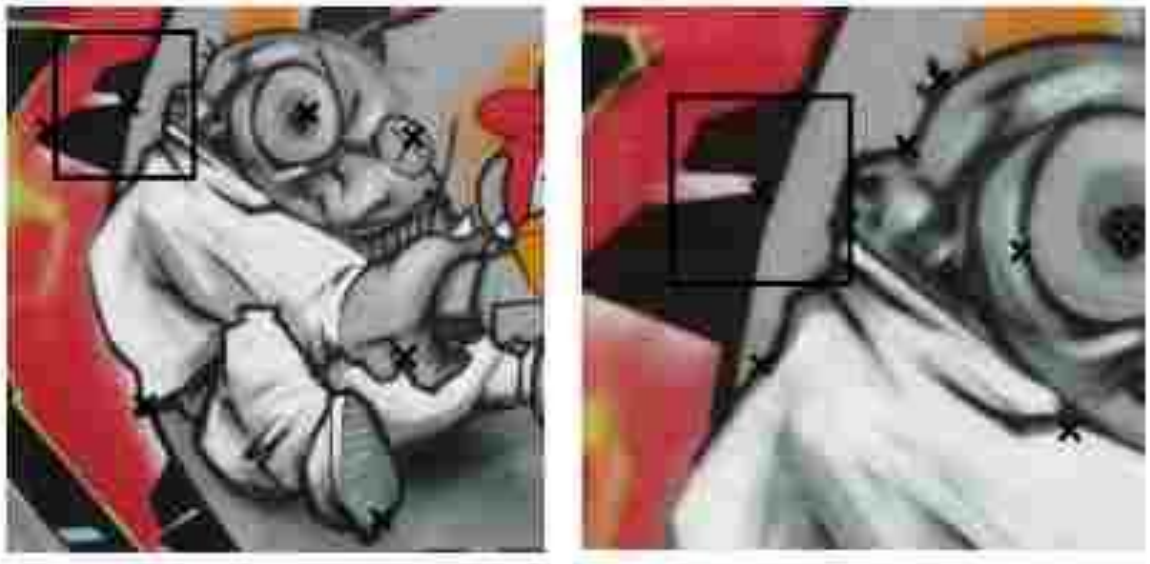
- Cum adaptam pentru schimbari de scala?





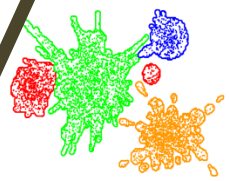


Metoda Harris multi-scala









Metoda Harris multi-scala



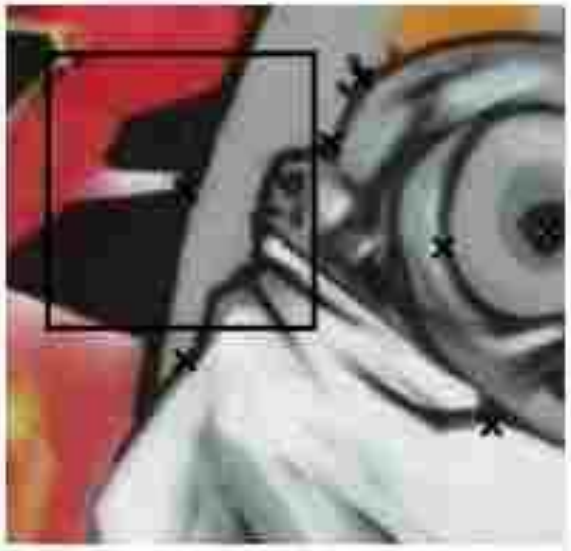






Metoda Harris multi-scala



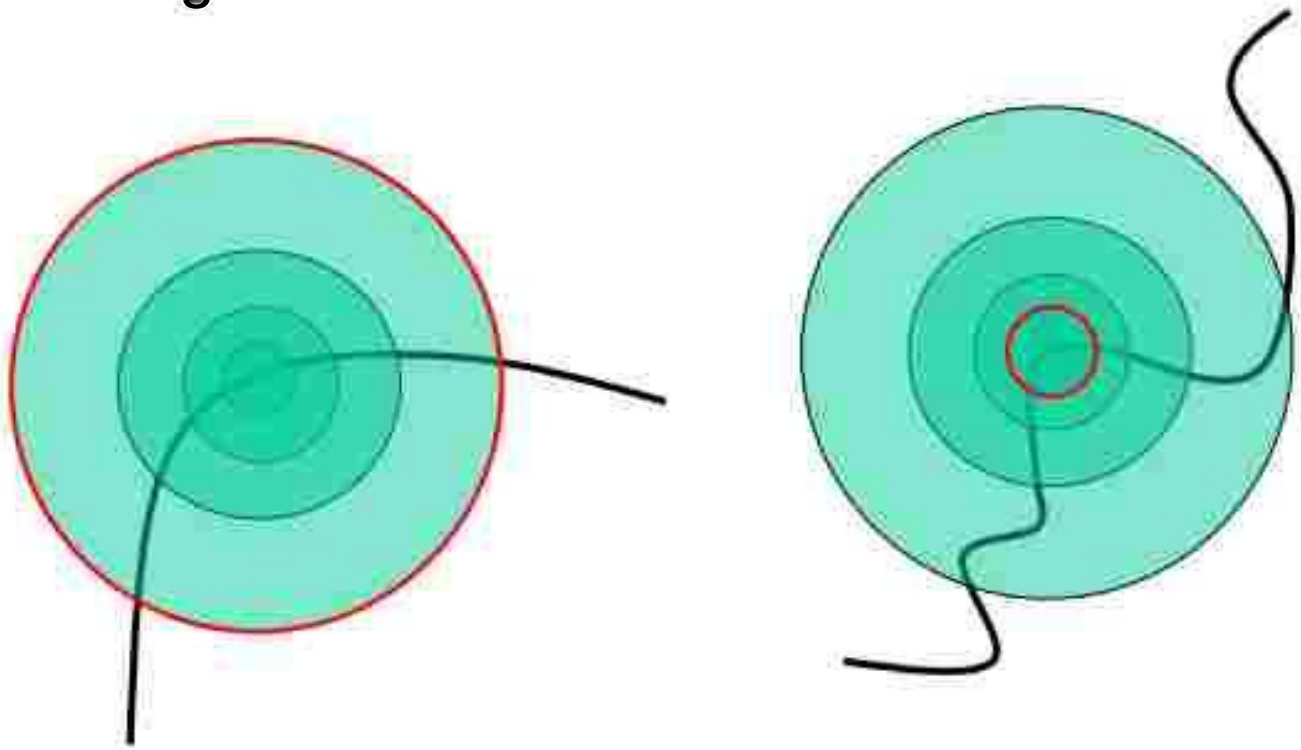






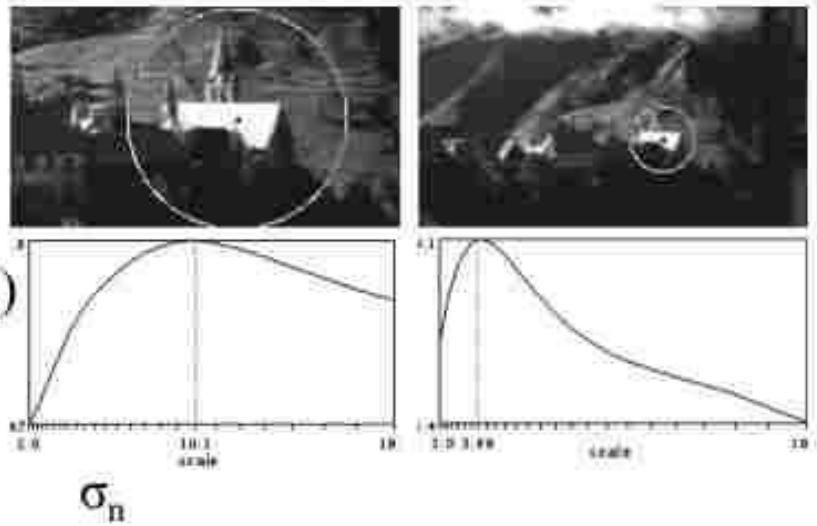
Metoda Harris multi-scala

Cum alegem scala?



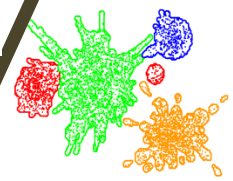
Metoda Harris multi-scala

- Design a function $F(x,\sigma_n)$ which provides some local measure.
- Select points at which $F(x,\sigma_n)$ is maximal over σ_n.



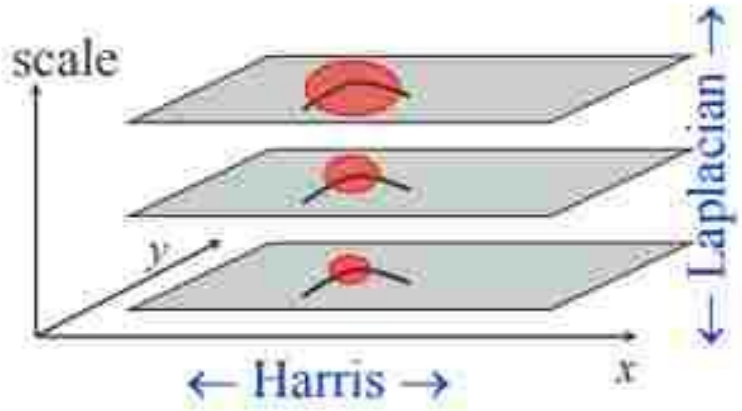
max of $F(x,\sigma_n)$ corresponds to characteristic scale!

T. Lindeberg, "Feature detection with automatic scale selection" *International Journal of Computer Vision*, vol. 30, no. 2, pp 77-116, 1998.

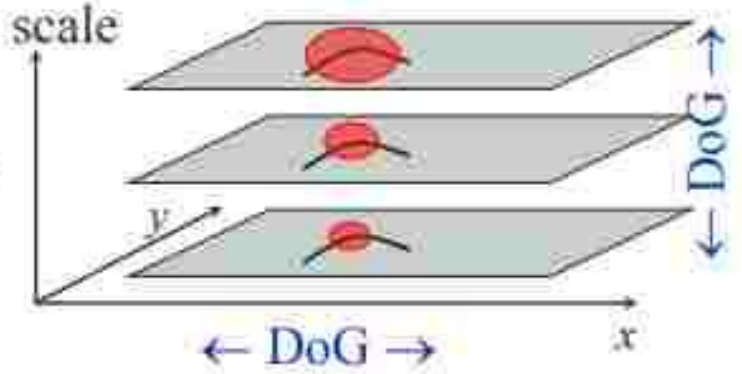


Metoda Harris multi-scala

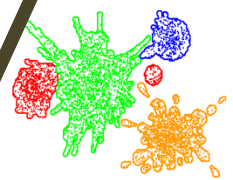
- Harris-Laplacian[1]
  Find local maximum of:
  - Harris corner detector in space (image coordinates)
  - Laplacian in scale



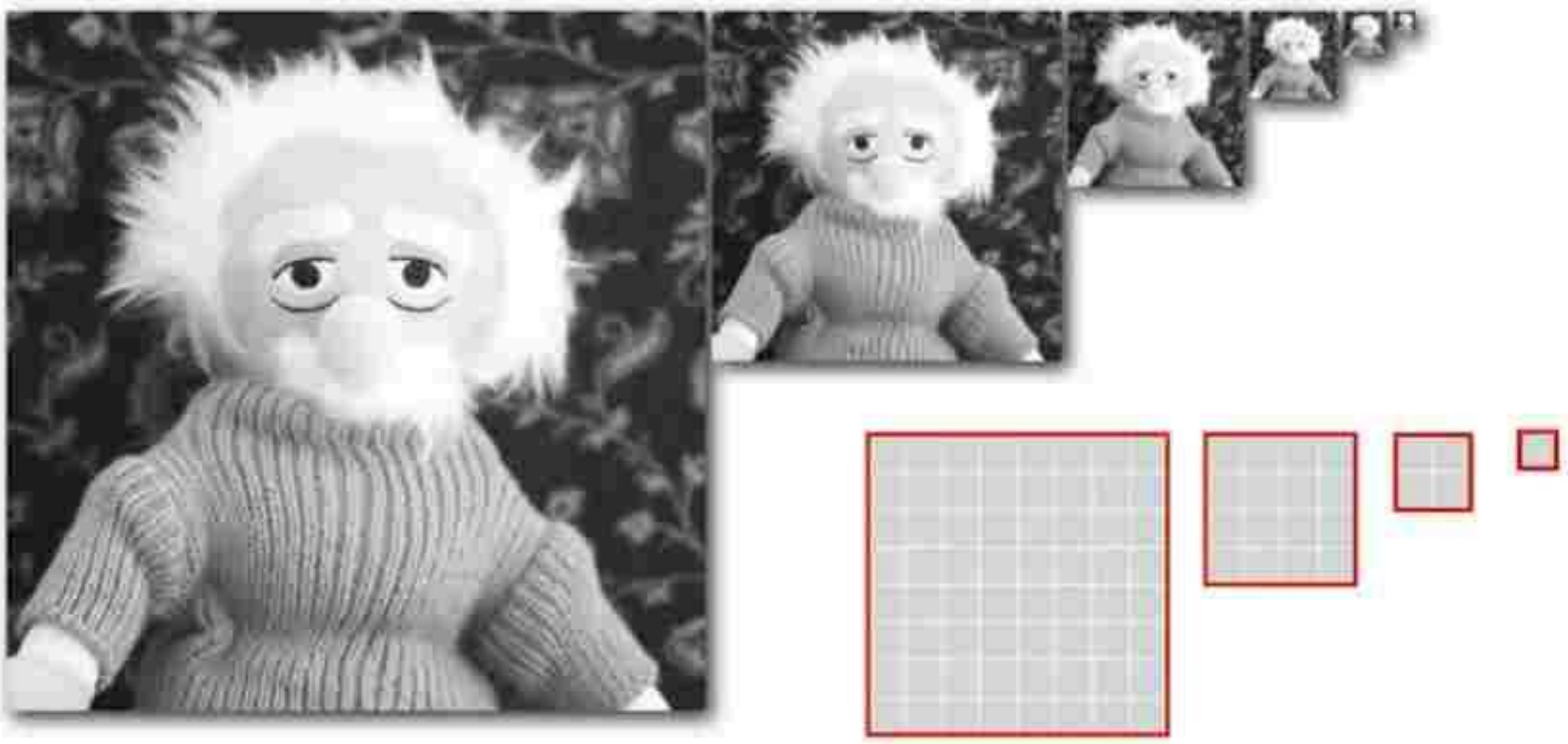
- SIFT (Lowe)[2]
  Find local maximum of:
  - Difference of Gaussians in space and scale



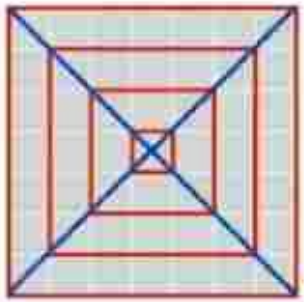
[1] K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004

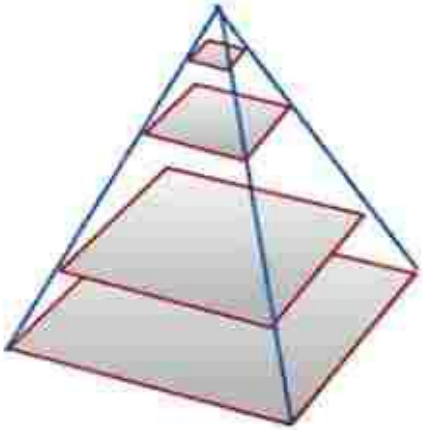


Piramide Gaussiene







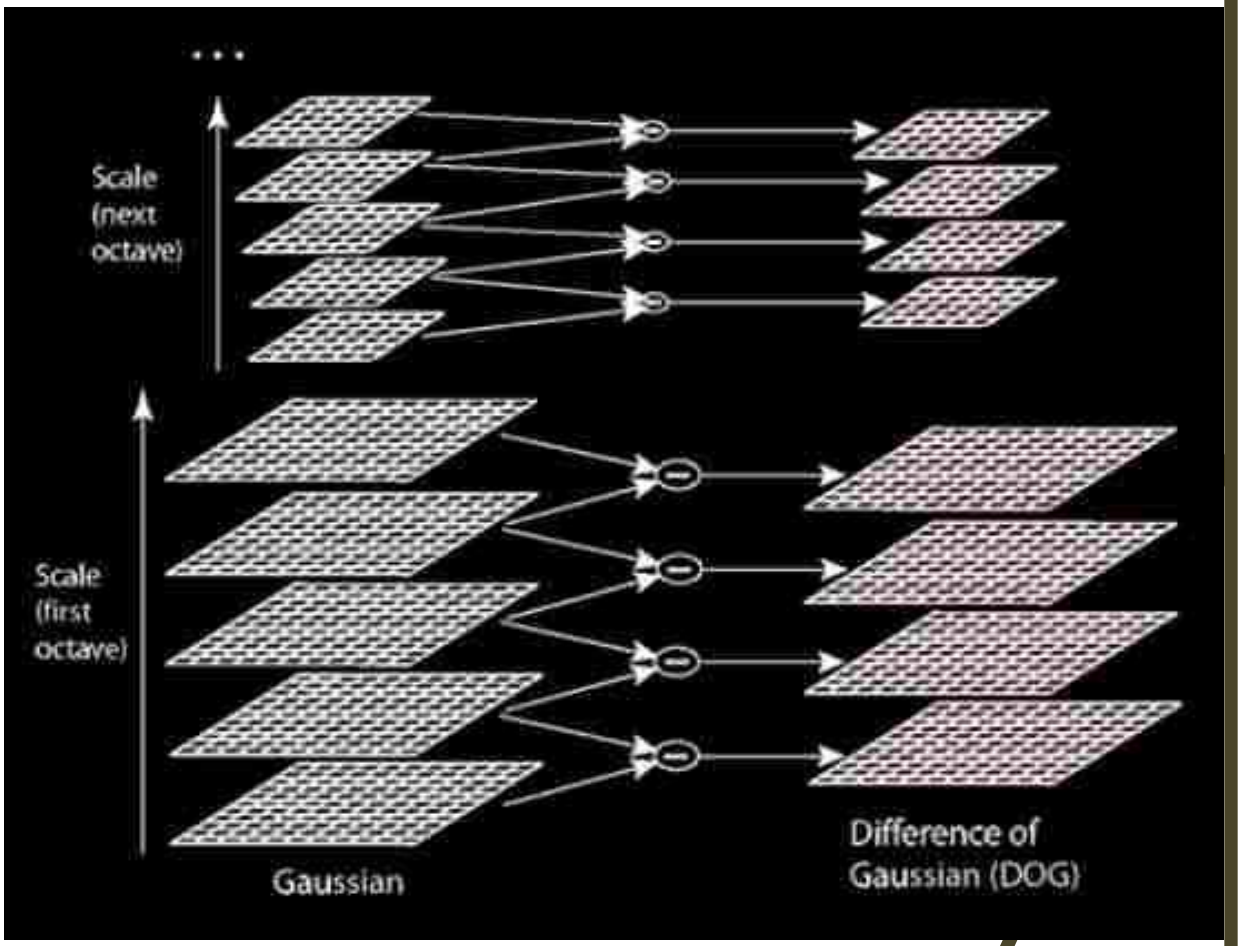


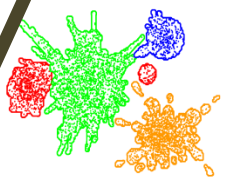


g_1 g_2 g_3 g_4 g_5

SIFT

1. Imaginea se filtreaza cu nuclee gaussiene cu varianta din ce in ce mai mare.
2. Se face diferenta intre imagini filtrate cu nuclee consecutive (DoG)
3. Se repeta pentru fiecare scala



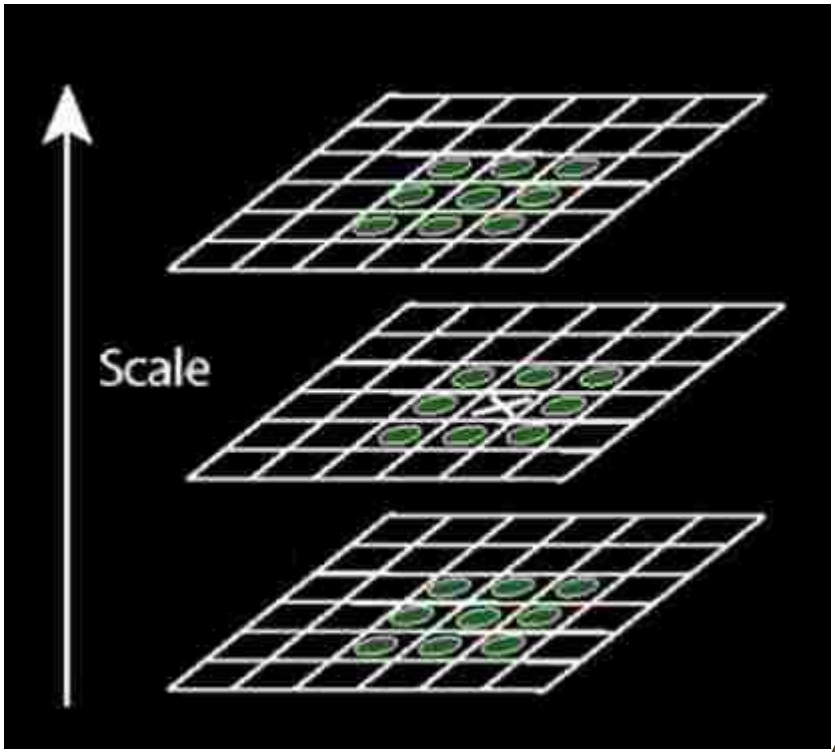


SIFT

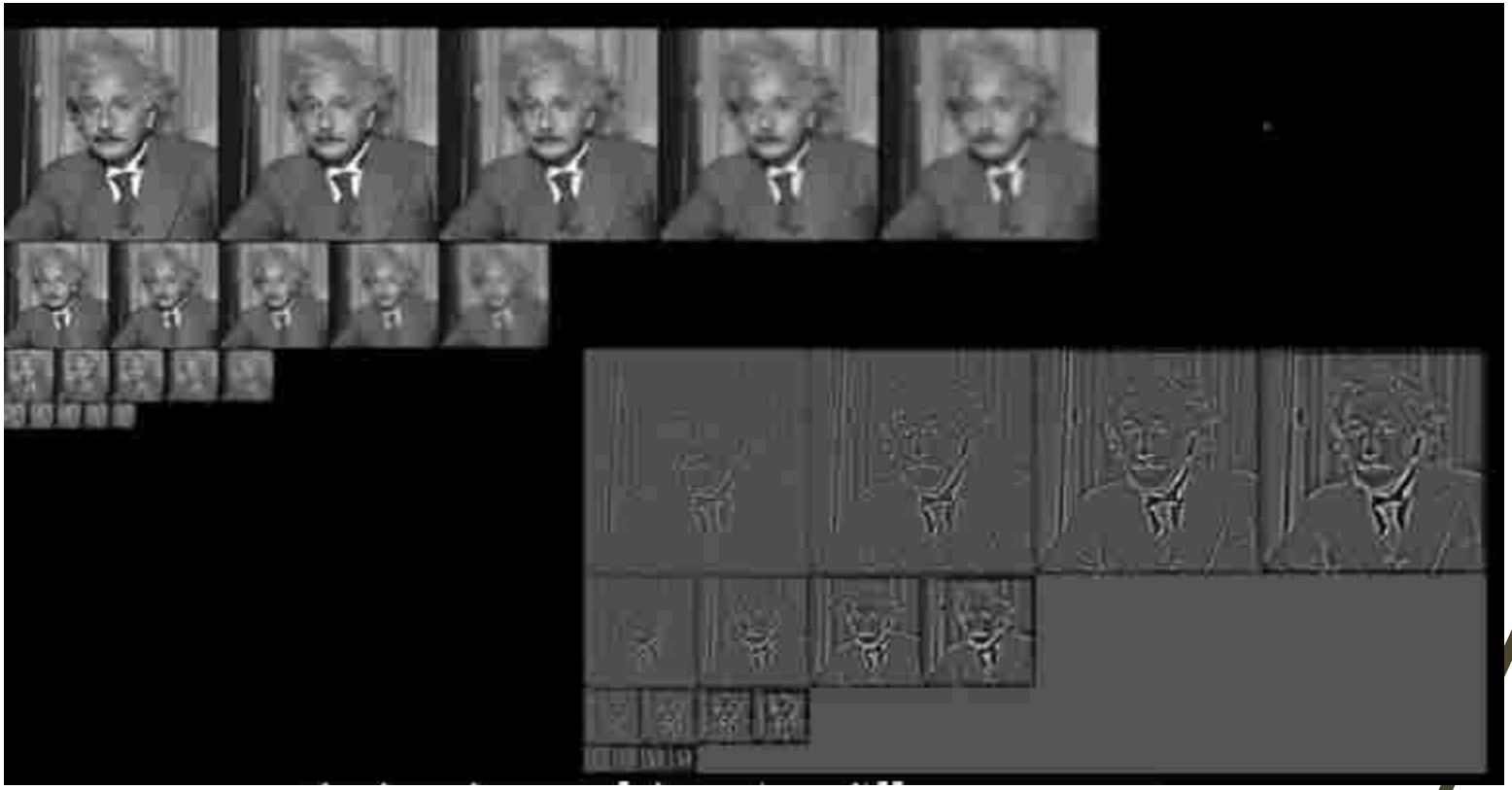
Se compara valoarea pixelului curent cu cei 26 vecini:

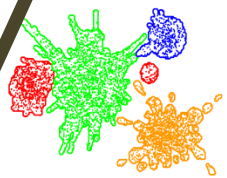
- 8 la aceeasi scala,
- 9 la scala urmatoare,
- 9 la scala precedenta

Se pastreaza daca e maxim

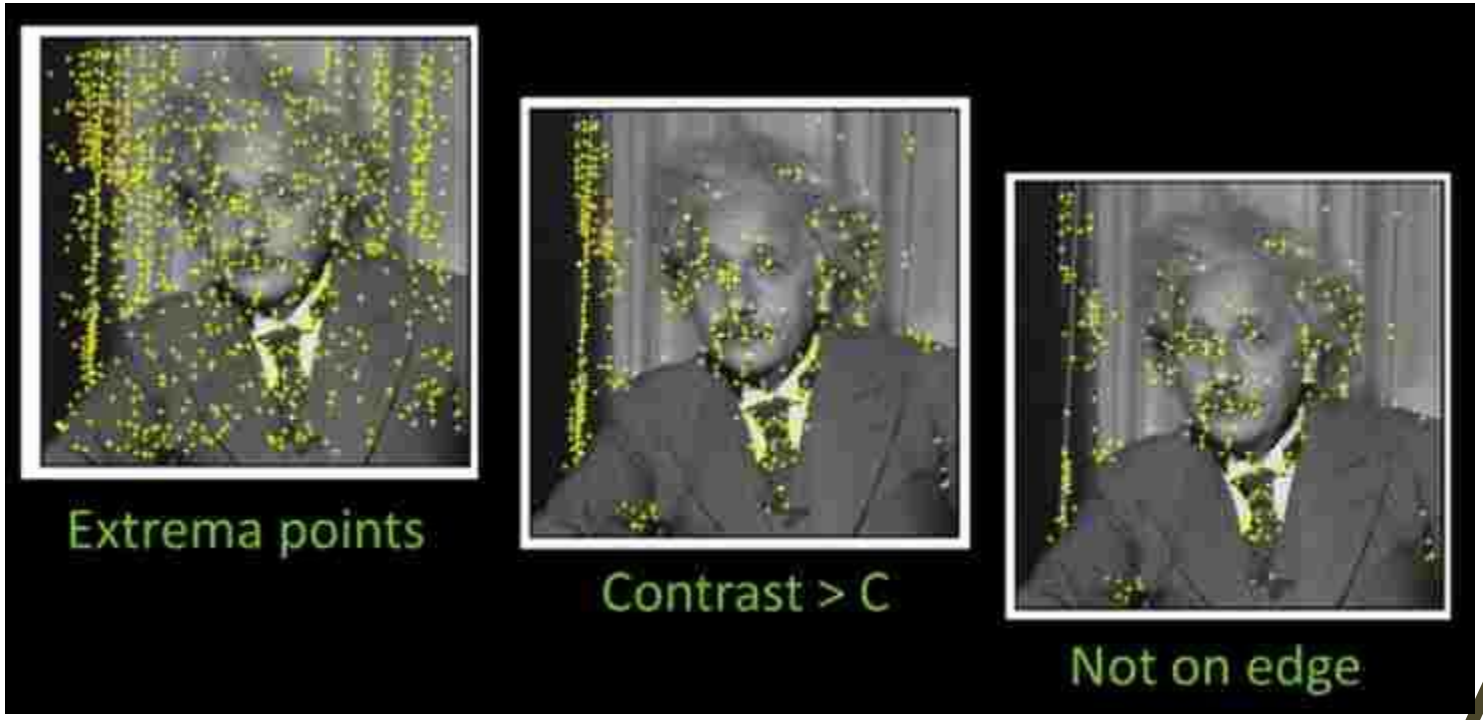


SIFT

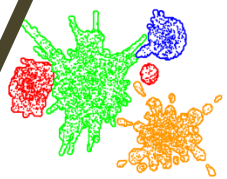




SIFT



Se elimina punctele de contur si cele cu contrast mic.



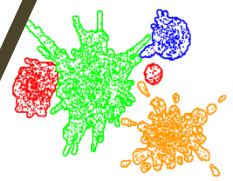
SIFT

Overall SIFT Procedure

- Scale-space extrema detection
- Keypoint localization

Or use Harris-Laplace or other method

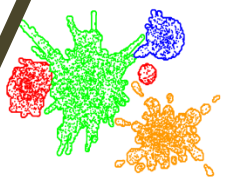
- Orientation assignment
- Keypoint description



SIFT

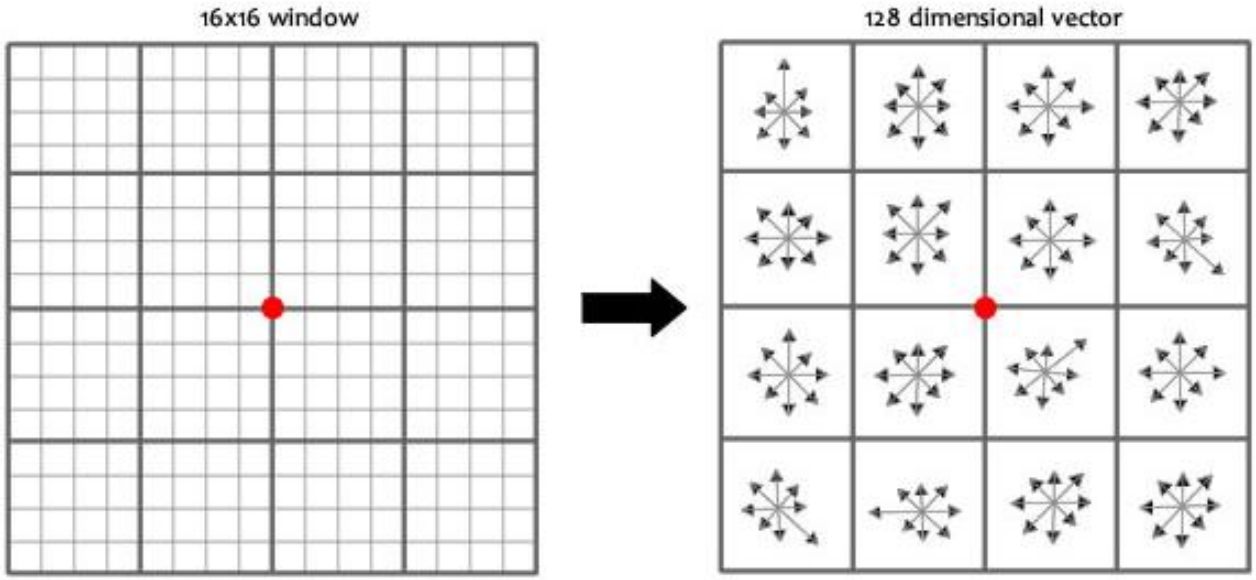
Alegerea orientarii

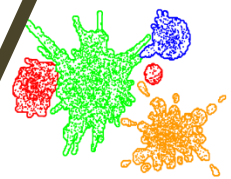
- Se defineste zona de interes:
  - 16x16 pixeli centrati in pixelul de interes la scala la care a fost gasit
- In zona punctului de interes se calculeaza HOG si se pastreaza orientarea predominanta
  - argmax(HOG)
- Se roteste zona de interes astfel incat orientarea predominanta sa devina ‘nord’
- Se scaleaza zona de interes la o rezolutie fixa



SIFT

- Construirea descriptorului
  - Se imparte zona de interes in 4x4 blocuri de 4x4px
  - Se calculeaza HOG pe fiecare bloc (8 binuri)
    - Imaginea de orientari e blurata gaussian
    - HOG-urile sunt normalizate
  - Fiecare punct e caracterizat de cele 16 histograme





Bibliografie

- Ionut Mironica – IVOM, note de curs
- Aaron Bobick, Irfan Essa, Arpan Chakraborty “Introduction to Computer Vision”, https://www.udacity.com/course/introduction-to-computer-vision--ud810
- Irfan Essa, David Joyner, Arpan Chakraborty “Computational Photography“, https://www.udacity.com/course/computational-photography--ud955
- William Hoff “Computer Vision” course https://www.youtube.com/playlist?list=PL4B3F8D4A5CAD8DA3

