

LAPI – Laboratorul de
Analiza și Prelucrarea
Imaginilor



Universitatea
POLITEHNICA din
București



Facultatea de Electronică,
Telecomunicații și
Tehnologia Informației

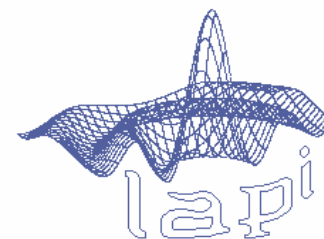
Interfață Vizuală Om-Mașină

Analiza și recunoașterea gesturilor

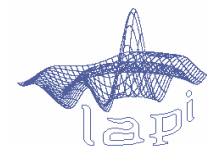
Dr.ing. Ionuț Mironică

<http://ionut.mironica.ro>

Prof. Dr. Ing. Bogdan Ionescu



LAPCM



LAPI – Laboratorul de
Analiza și Prelucrarea
Imaginilor



Universitatea
POLITEHNICA din
București



Facultatea de Electronică,
Telecomunicații și
Tehnologia Informației

Materiale de curs

Cursurile pot fi descărcate de la adresa:

<http://ionut.mironica.ro/teaching.htm>

ionut.mironica [at] gmail.com

Laborator B138

Examen final (scris): 50%

Cuprins curs

- Noțiuni introductive
- Dispozitive pentru detecția gesturilor
- Algoritmi de recunoaștere automată a gesturilor
 - Algoritmi de detecție după formă
 - Algoritmi pentru urmărirea traiectoriei
 - Detecția mișcării
- Algoritmi de clasificare

Cuprins curs 1

- Noțiuni introductive
- Domenii de utilizare
- Dispozitive de recunoaștere a gesturilor
 - Dispozitive contact / non-contact
 - Camere de adâncime
 - Consola Kinect
 - Testare Kinect
 - Aplicații Kinect
- Concluzii

Bibliografie

- [1] „The Human-Computer Interaction Handbook, Fundamentals, Evolving, Technologies and Emerging Applications”, Ed. A. Sears, J.A. Jako, LEA – Taylor & Francis Group, New York – London, 2008;
- [2] „Computer Vision: Algorithms and Applications”, Rick Szeliski (draft online);
- [3] „Computer Vision : A Modern Approach”, David A. Forsyth și Jean Ponce (draft online);
- [4] „Visual Object Recognition”, Kristen Grauman și Bastian Leibe;
- [5] „Computer Vision”, Linda G. Shapiro și George C. Stockman;
- [6] „Introductory Techniques for 3D Computer Vision”, E. Trucco și A. Verri;
- [7] „Pattern classification”, Richard O. Duda, Peter E. Hart și David G. Stork;
- [8] „Pattern Recognition and Machine Learning”, Christopher M. Bishop;
- [9] „Conceptul de Indexare Automată după Conținut în Contextul Datelor Multimedia”, B. Ionescu, I. Mironică, Editura MartrixRom, 2013 (draft online).

I. Noțiuni introductive

Tendința actuală a sistemelor de interfațare om-mașină este de înlocuire a dispozitivelor mecanice:



I. Noțiuni introductive

- Una dintre cele mai populare direcții de studiu o constituie analiza și recunoașterea gesturilor (expresive, semnificative și informaționale);
- Recunoașterea de gesturi permite comunicarea om-mașină în mod natural fără a interacționa cu alte dispozitive mecanice, deci fără a fi nevoie de a învăța regului de utilizare a acestora.
- În acest mod, dispozitivele clasice (mouse, touch-screen, tastatură) pot devini redundante.

I. Noțiuni introductive

Aplicații



Medicale:

- ajutor pentru persoanele cu dificultăți de auz precum recunoașterea automată a limbajului semnelor;
- monitorizarea pacienților la nivel emoțional, stres;
- chirurgie.

Criminalistică:

- tehnici de identificare;
- detector de minciuni.

I. Noțiuni introductive

Aplicații



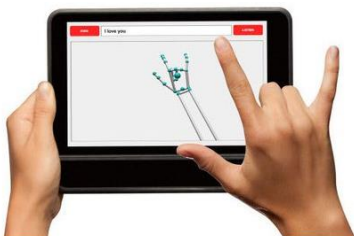
“Entertainment”:

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



Industrie:

- manipulare echipamente / roboți industriali.



Asistare persoane cu dizabilități:

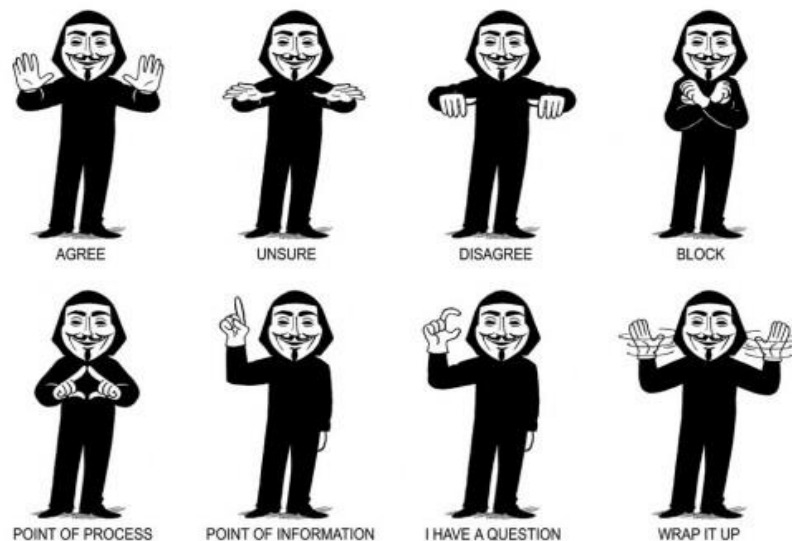
ex: tabletă cu recunoaștere automată a semnelor.

I. Noțiuni introductive

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;



I. Noțiuni introductive

Tipuri de gesturi



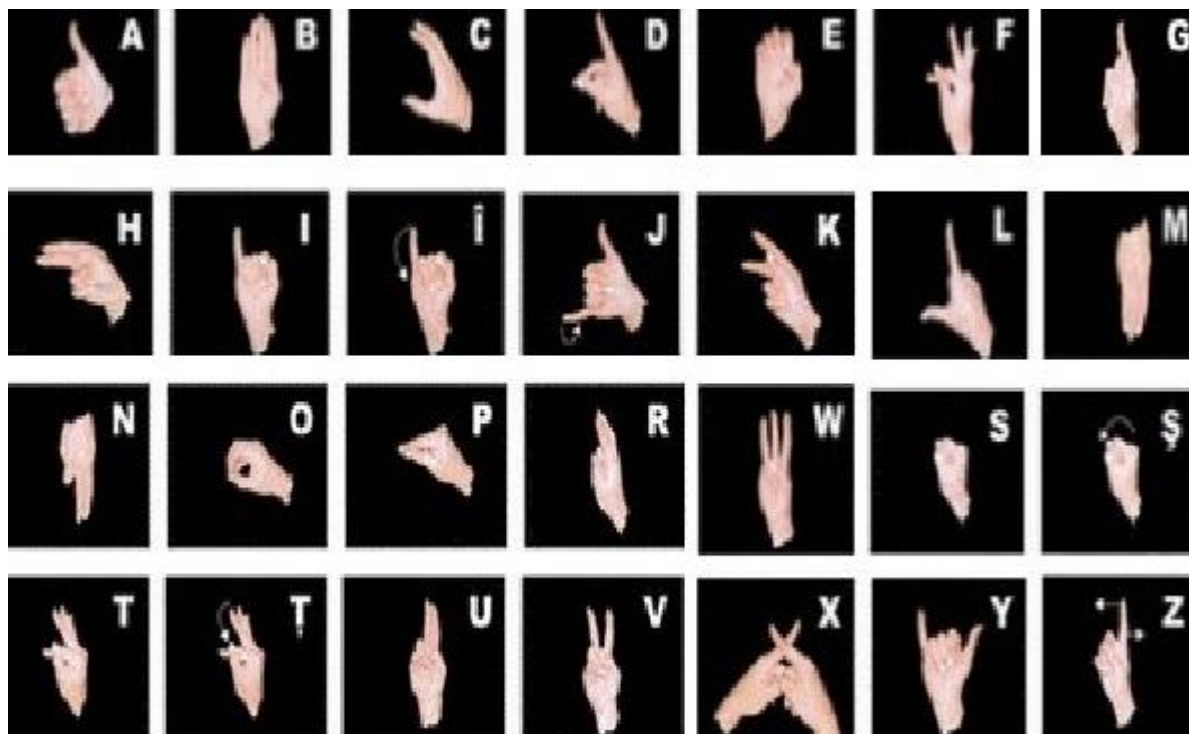
American Sign Language

- Comunicare non-verbală – mod de lucru pentru peste 500.000 de persoane din America de Nord;
- Cuvintele aditionale se formează prin utilizarea alfabetului;
- Conțin și gesturi dinamice (“J” și “Z”).

I. Noțiuni introductive

Tipuri de gesturi

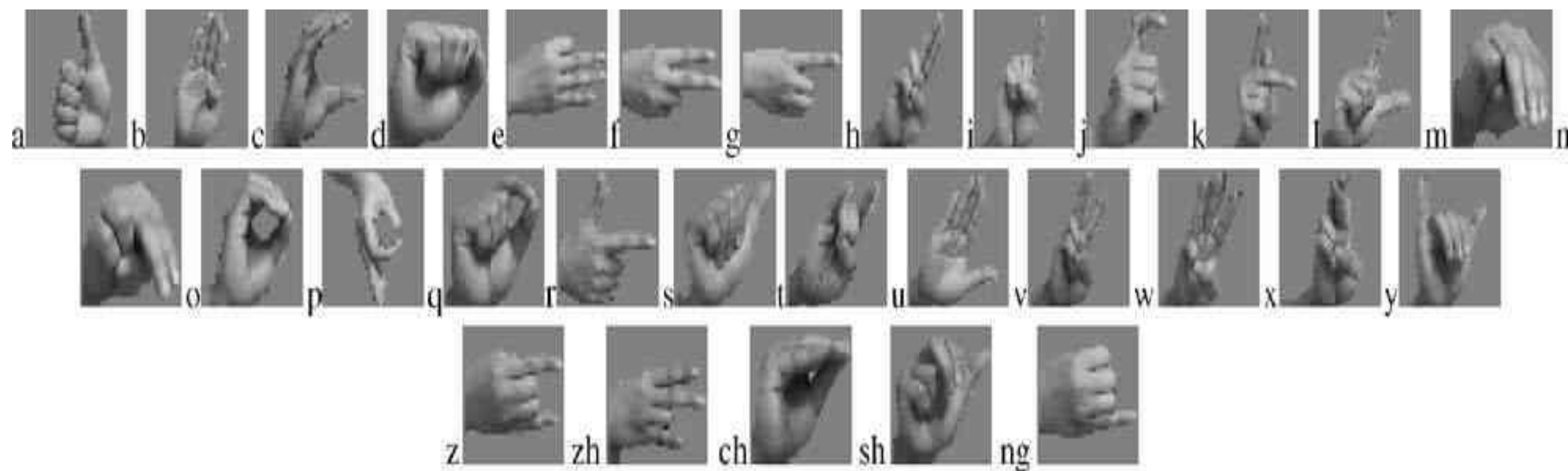
Limba mimico-gestuală românească



I. Noțiuni introductive

Tipuri de gesturi

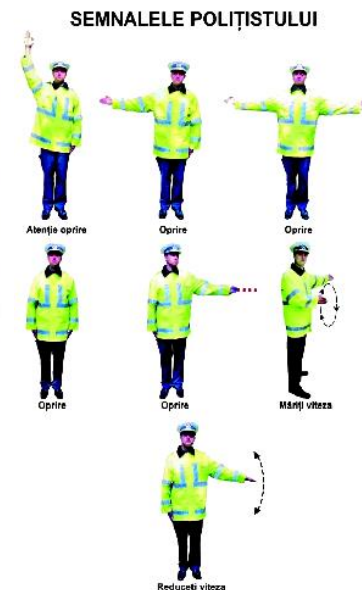
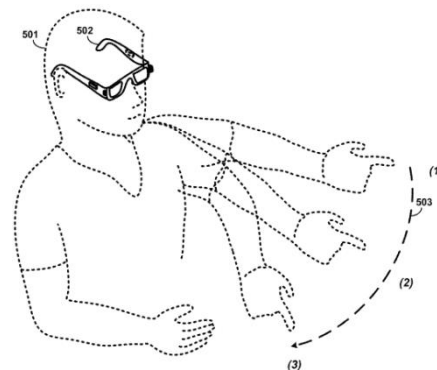
Limba mimico-gestual chinezesc



I. Noțiuni introductive

Tipuri de gesturi

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



I. Noțiuni introductive

Tipuri de gesturi

Gesturi ale capului și feței: exemple, a da din cap, direcția privirii, ridicarea sprâncenelor, deschiderea gurii pentru a vorbi, clipire, sau expresii ale feței precum: neutră, surpriză, dezgust, bucurie, frică, supărare, furie etc.;



I. Noțiuni introductive

Tipuri de gesturi

Gesturi ale corpului: implică mișcarea întregului corp, precum urmărirea mișcării a două persoane ce interacționează, mișcările unui dansator, analiza cinematicii corpului - antrenare;

- 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;

I. Noțiuni introductive

Tipuri de gesturi

Gesturi ale corpului: implică mișcarea întregului corp, precum urmărirea mișcării a două persoane ce interacționează, mișcările unui dansator, analiza cinematicii corpului - antrenare;

- Supraveghere automată
- Detecție automată a evenimentelor



I. Noțiuni introductive

Tipuri de gesturi

Blip10000



Art



Auto



Comedy



Food



Music

UCF101



Answer phone



Chop banana



Dial phone



Drink water



Eat banana

Activities of Daily Living dataset



Basketball



Bench press



Billiards



Diving



Drumming

Mediaeval 2013 - Violent Scenes Dataset



Billy Elliot



I am legend



Saving Private Ryan



The Sixth Sense

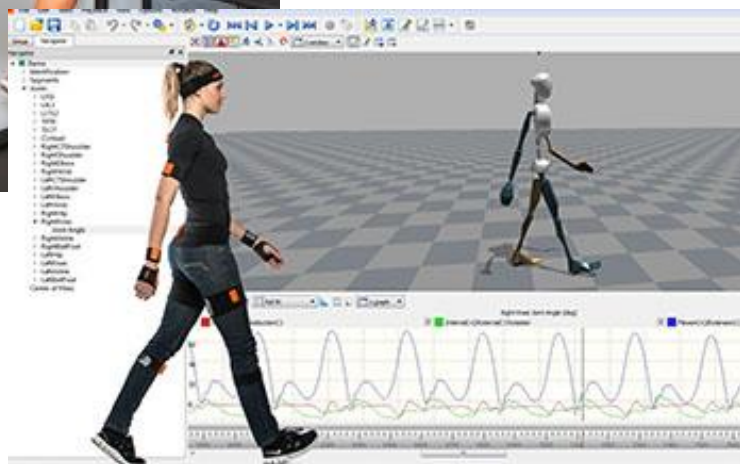
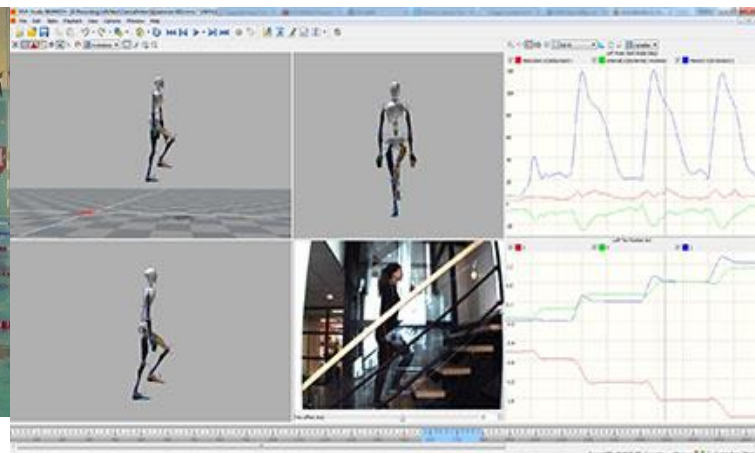
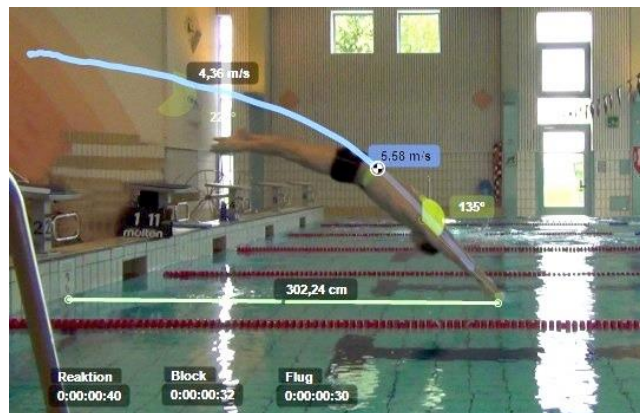


Pirates Of The Caribbean

I. Noțiuni introductive

Tipuri de gesturi

Îmbunătățirea performanțelor sportive



[www.xsens.com]

I. Noțiuni introductive

Provocări

- 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).

I. Noțiuni introductive

Provocări

Sindromul „Gorilla arm”

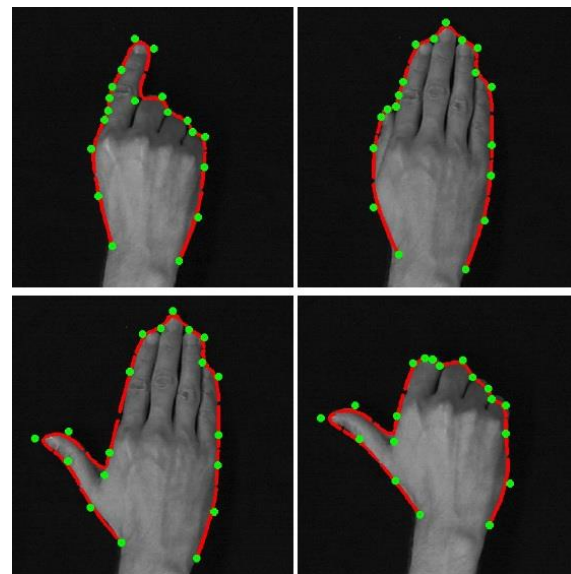
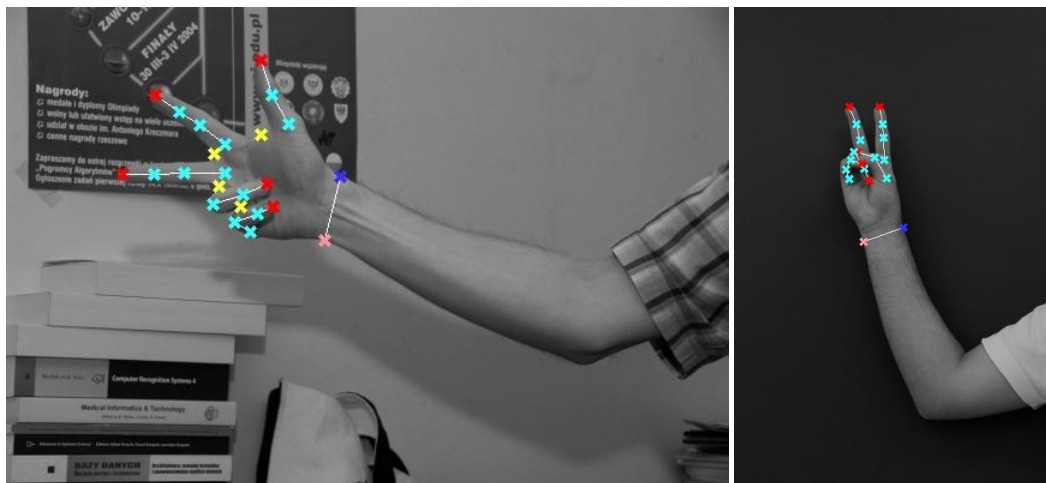
- După o utilizare îndelungată a mâinilor poate apare sindromul „Gorilla arm”;
- Oamenii nu pot „ține” mâinile în fața corpului pentru o perioadă lungă de timp deoarece apar dureri, crampe și oferă impresia de mâine „supradimensionate”;
- Acesta a dus de altfel și la dispariția tehnologiei touch-screen în anii 1980;
- În domeniu există o vorbă “Remember the gorilla arm!” ceea ce reprezintă o aluzie la “How is this going to fly in *real* use?”.



II. Tehnici de prelucrare gesturi

Clasificare

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

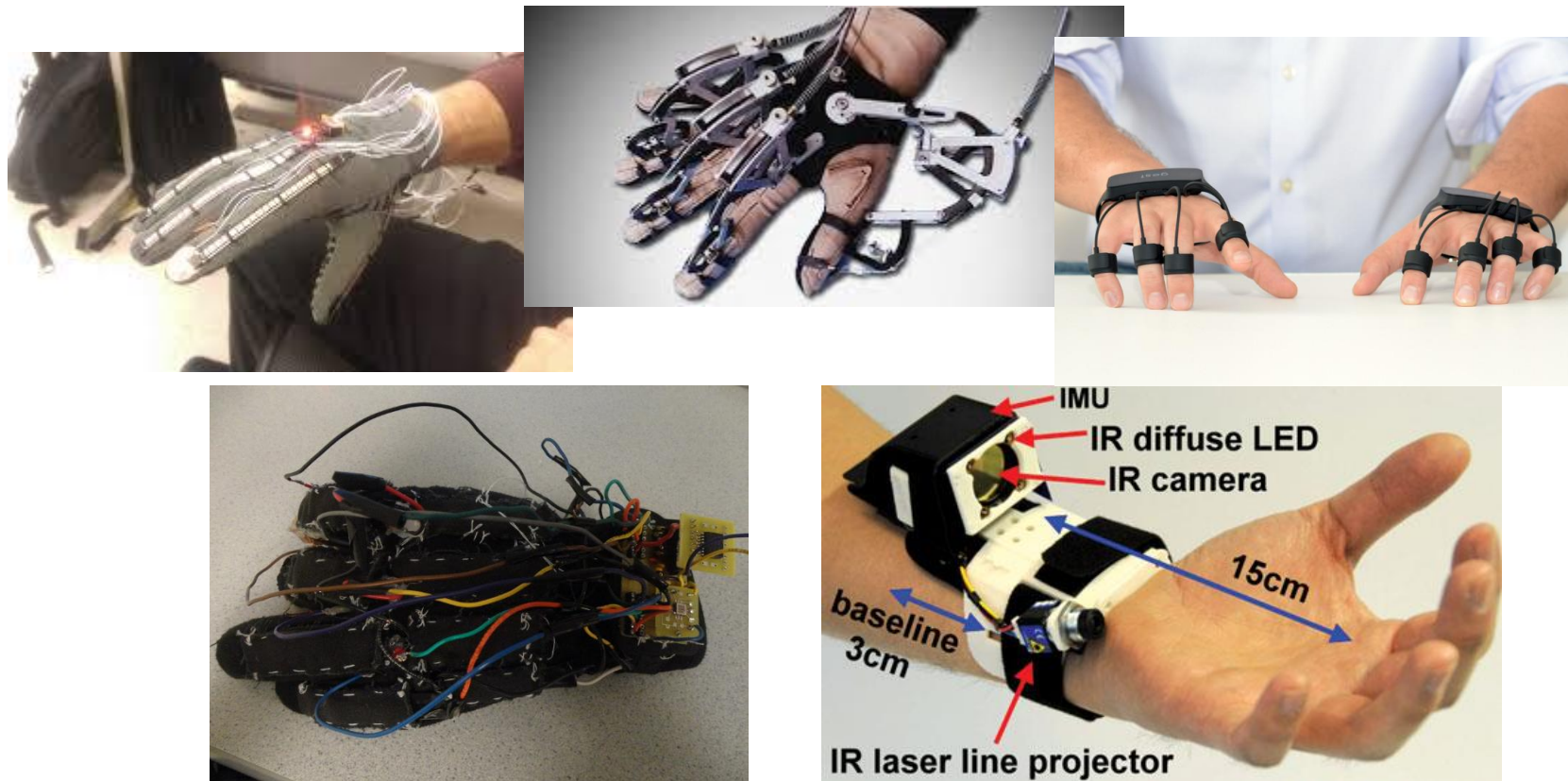


[<http://sun.aei.polsl.pl/~mkawulok/gestures/>]

II. Tehnici de prelucrare gesturi

Clasificare

- **(cu contact)** bazate pe date preluate de către dispozitive hardware specifice.



II. Tehnici de prelucrare gesturi

Companii



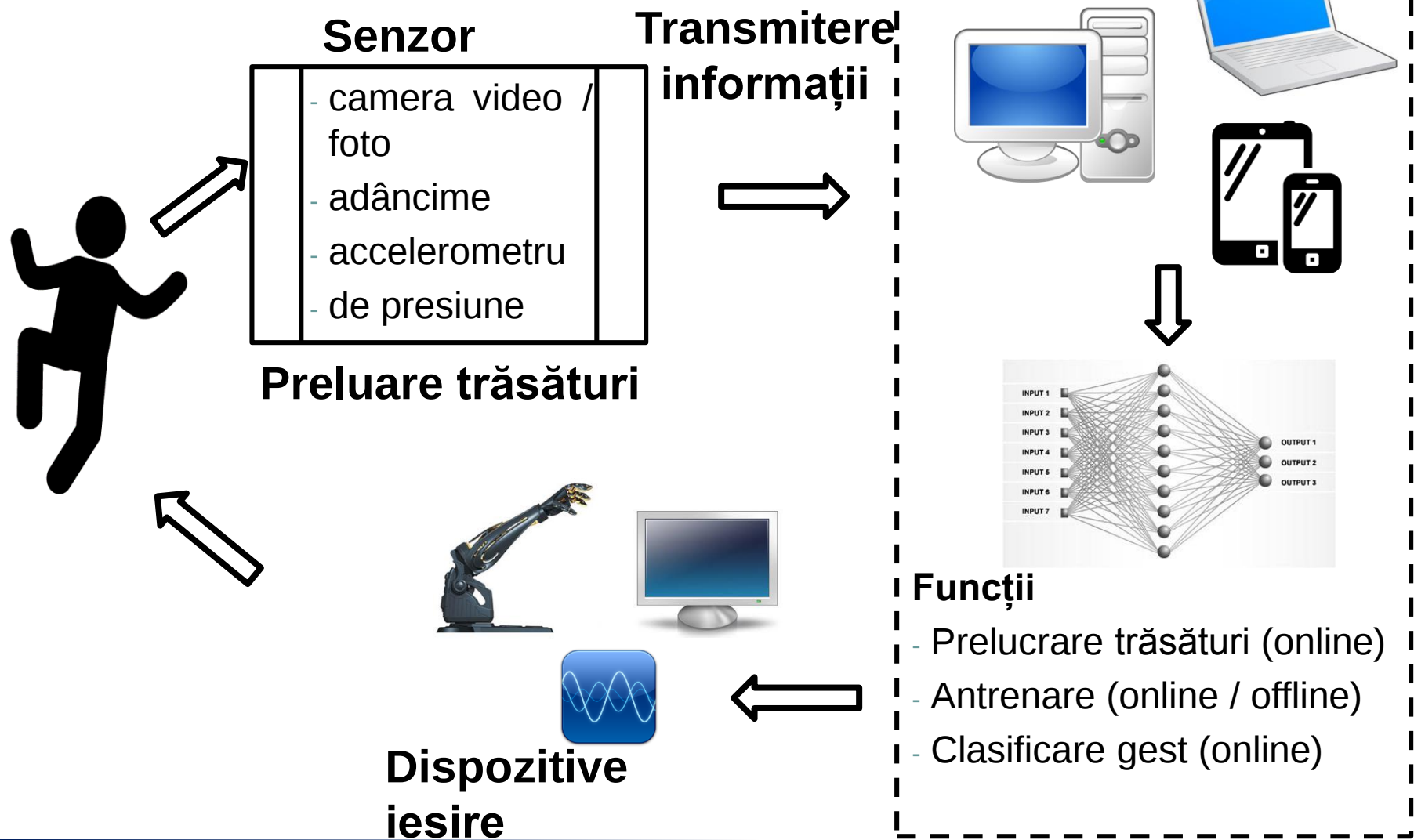
II. Tehnici de prelucrare gesturi

Companii



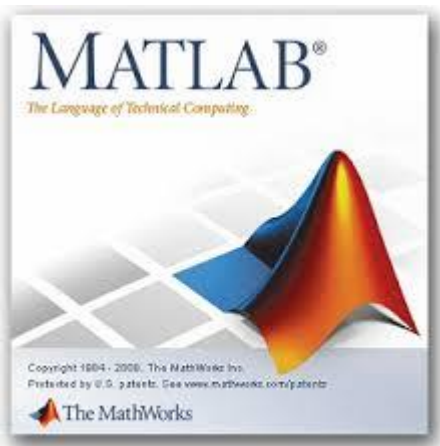
II. Tehnici de prelucrare gesturi

Schemă generală



II. Tehnici de prelucrare gesturi

Biblioteci / Limbaje pentru IVOM



- Computer Vision Toolbox;
- Neural Network Toolbox;
- Bioinformatics Toolbox (Support Vector Machines);
- Statistics Toolbox (Hierarchical Clustering, KMeans, Gaussian Mixture Models, Naive Bayes, Discriminant Analysis, Nearest Neighbors, Classification Trees, Ensemble Classifiers, etc);

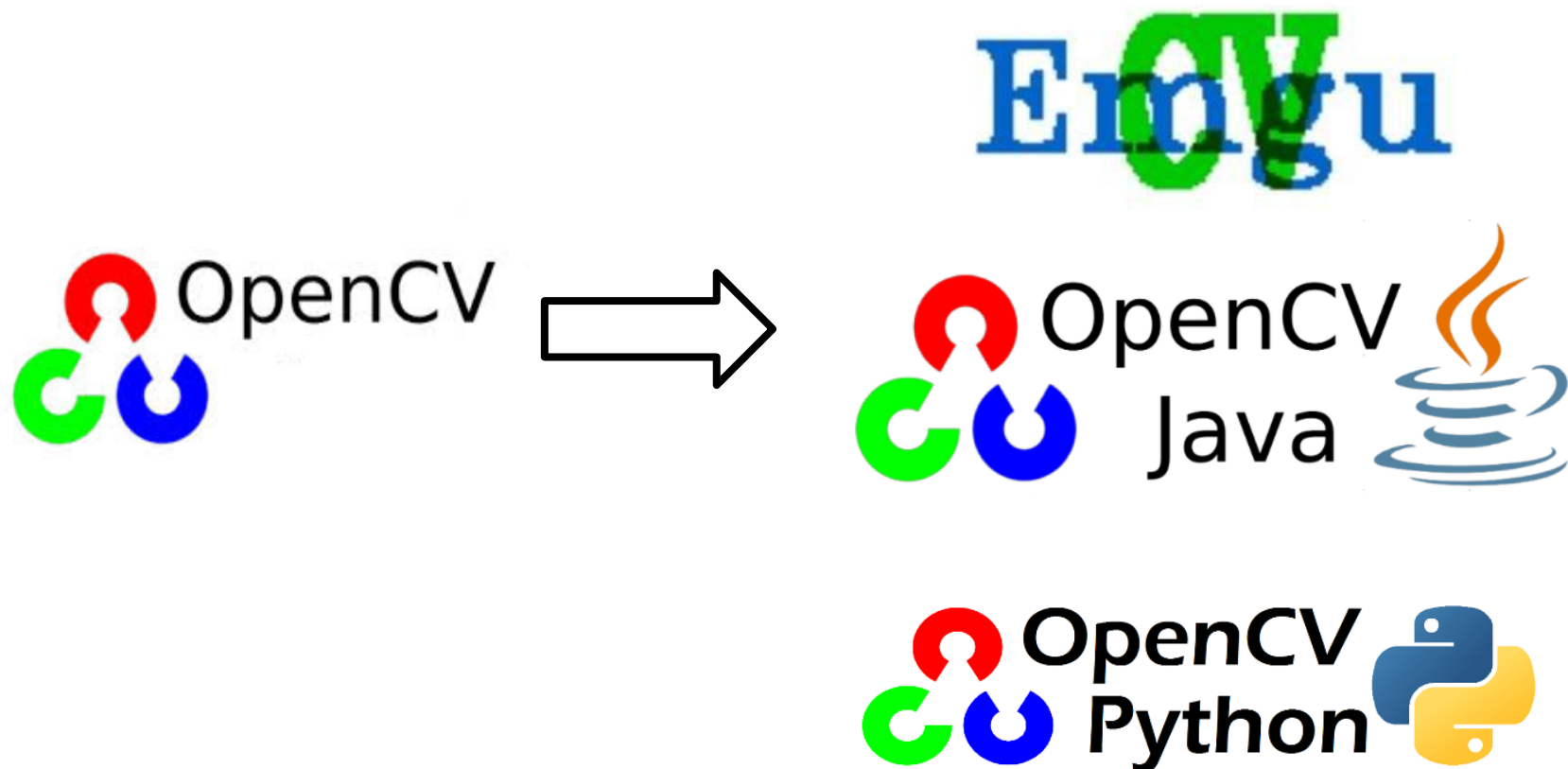
- platformă generală de computer vision ce include și facilități de clasificare, "open source", în C++;
- Machine Learning Library (MLL): Statistical Models, K-Nearest Neighbors, Support Vector Machines, Decision Trees, Boosting, Expectation Maximization, Neural Networks;
- Conține algoritmi de detecție a mișcării / detecție de puncte de interes (SIFT SURF, Harris, MSER etc)



II. Tehnici de prelucrare gesturi

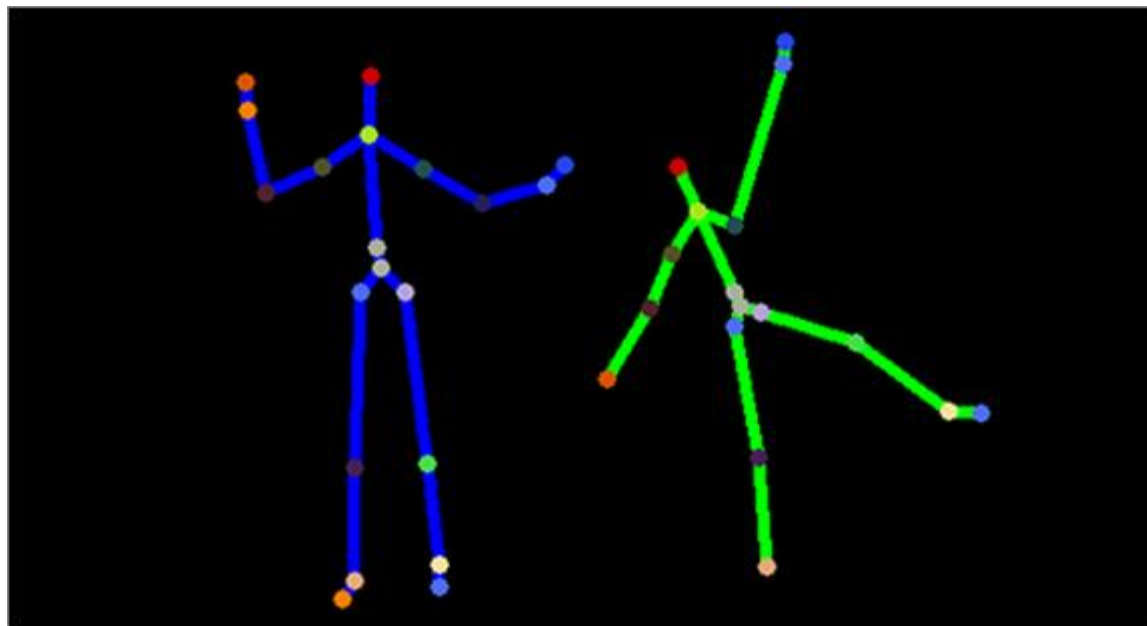
Biblioteci / Limbaje pentru IVOM

- Wrapper pentru diferite limbaje / sisteme (.NET / Android / Python)



II. Tehnici de prelucrare gesturi

Biblioteci / Limbaje pentru IVOM



II. Tehnici de prelucrare gesturi

Biblioteci / Limbaje pentru IVOM



II. Tehnici de prelucrare gesturi

Biblioteci / Limbaje pentru IVOM

- Halcon
- Matrox Imaging Library
- Open eVision
- Adaptive Vision Library
- Common Vision Blox

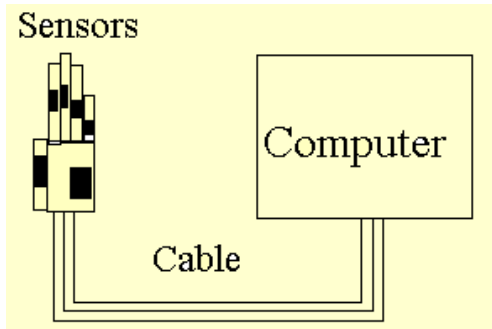
III. Dispozitive de recunoaștere a gesturilor

Dispozitive atașate

Mănuși

Tehnologie de captare a mișcărilor (ex îndoirea degetelor).

- Conțin un detector de mișcare (magnetic sau inerțial) care captează poziția globală a degetelor;
- Au un cost de producție destul de ridicat;
- Sunt folosite în special pentru manipulare în medii virtuale și de “realitate augmentată”.



III. Dispozitive de recunoaștere a gesturilor

Dispozitive atașate

Mănuși

- Datele sunt transmise wireless către un server calculator care le prelucrează în timp real;
- Avantaje:
 - nu mai sunt probleme de iluminare / ocluziuni.
- Dezavantaje:
 - preț ridicat.



III. Dispozitive de recunoaștere a gesturilor

Dispozitive atașate

Mănuși „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).



[MIT 2010]

III. Dispozitive de recunoaștere a gesturilor

Dispozitive atașate

Wii (Nintendo)



PS3 Move

Myo



III. Dispozitive de recunoaștere a gesturilor

Dispozitive atașate



Fitbit



WristQue



Cicret

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime

- Calculează distanța dintre cameră și obiectele din cameră;
- Majoritatea calculează distanța prin măsurarea distanței pe care lumina o parcurge de la senzor la obiect.



III. Dispozitive de recunoaștere a gesturilor

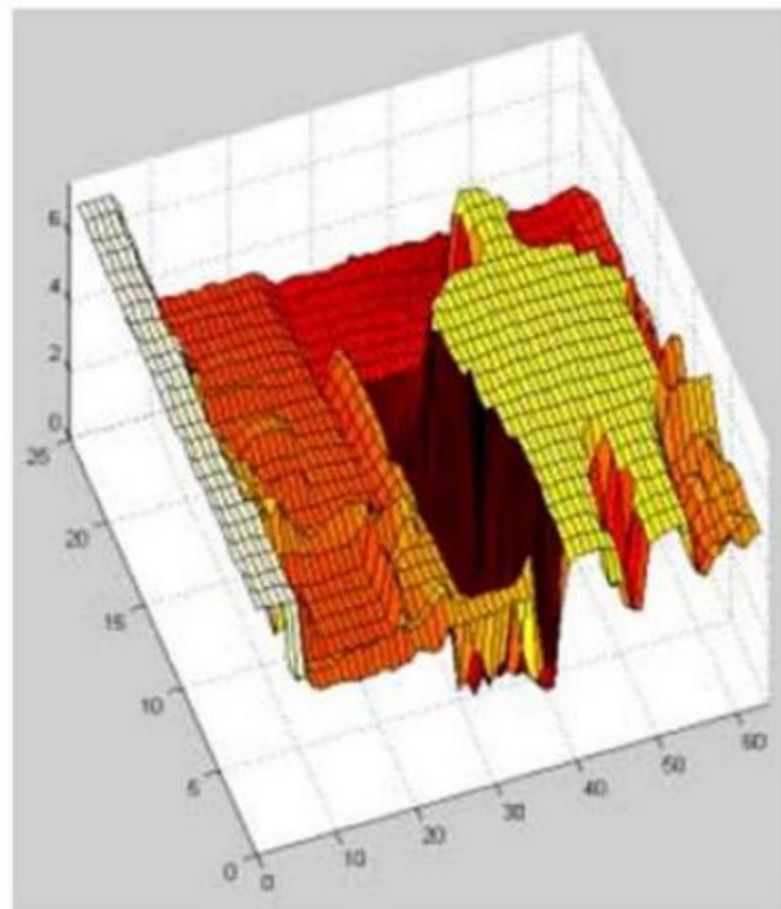
Camere de adâncime

- Ce afișează o cameră de adâncime?



III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime



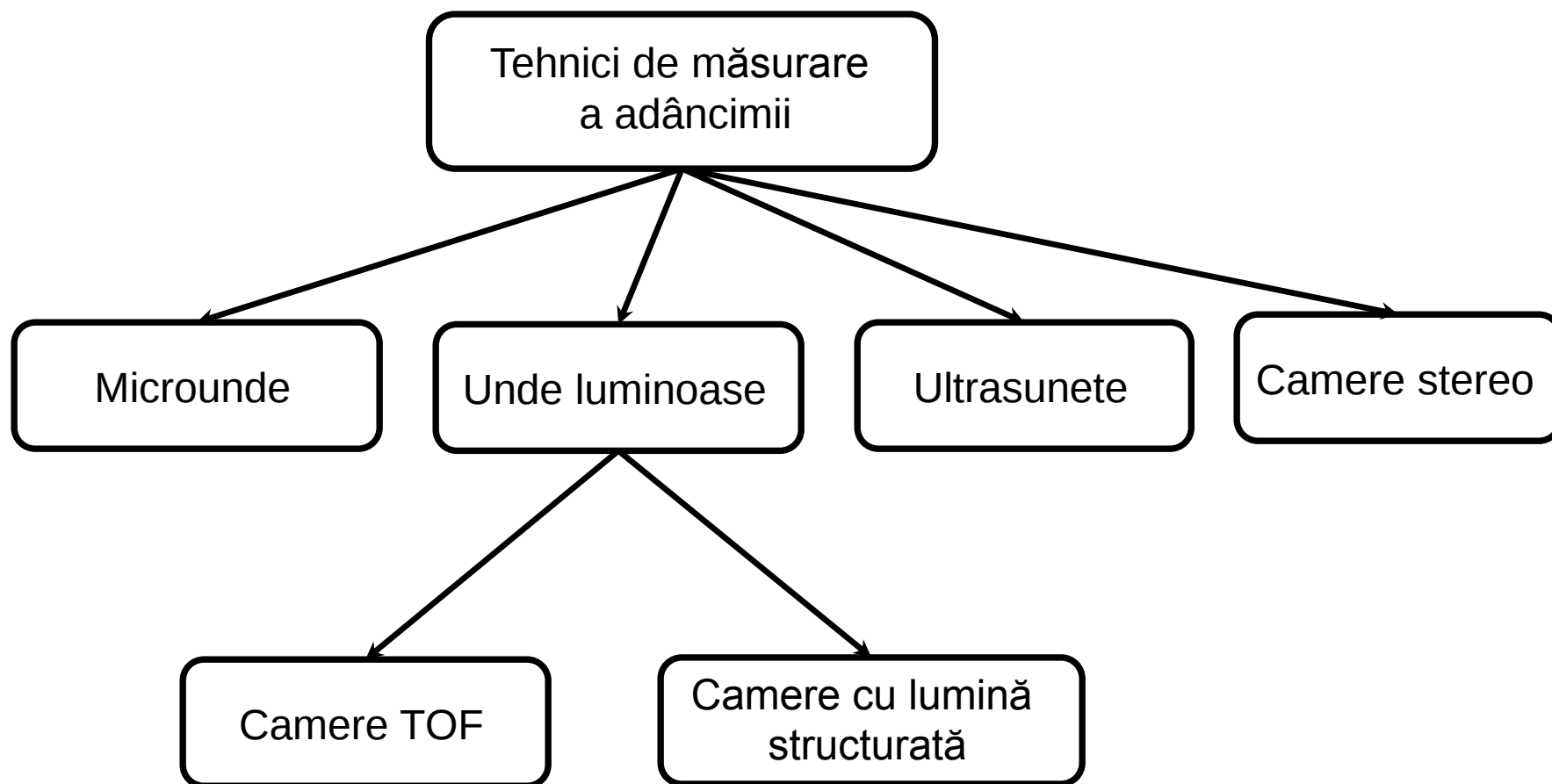
III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime



III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime



III. Dispozitive de recunoaștere a gesturilor

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



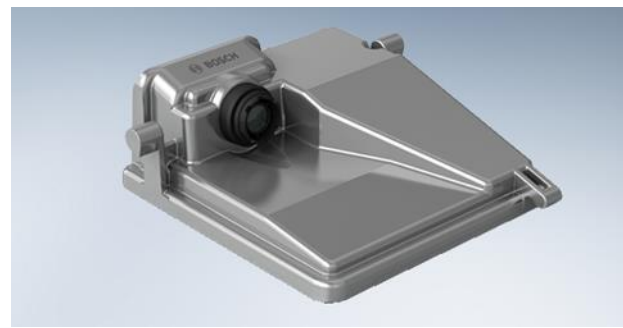
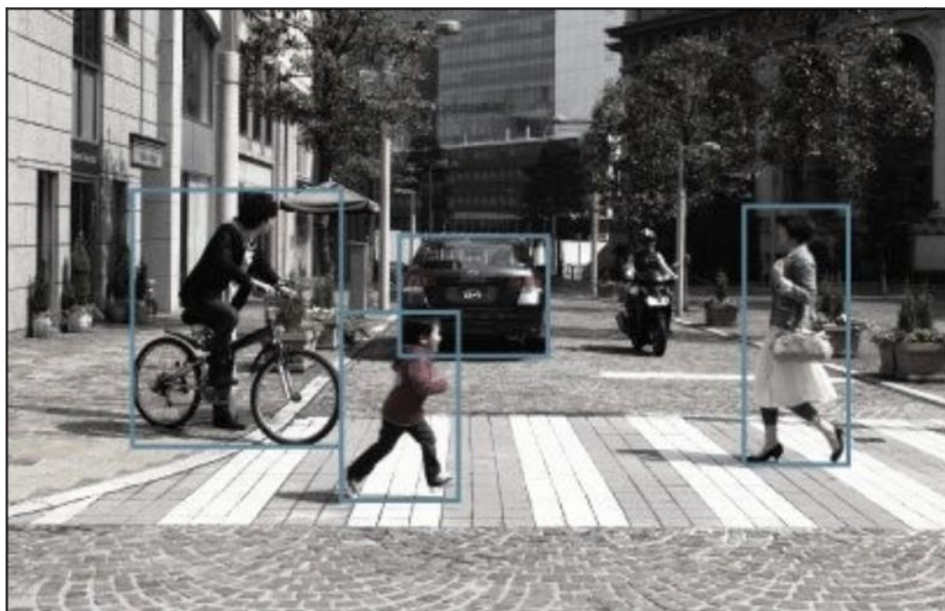
Camera Verascope

III. Dispozitive de recunoaștere a gesturilor

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).



III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Utilizare

- Pentru navigarea roboților;



Ex: robotul Nomad caută meteoriți în Antartica

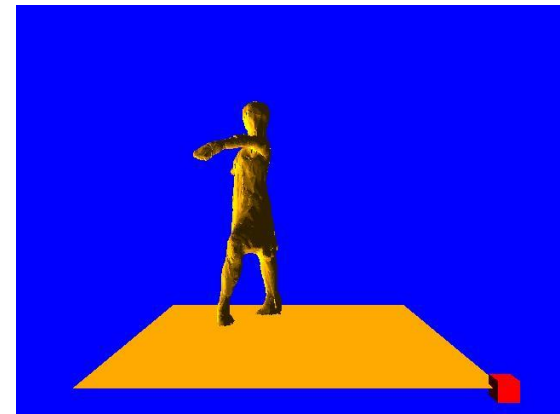
III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Utilizare

- Reconstrucție 3D;

Colectează flux video din 50 de camere

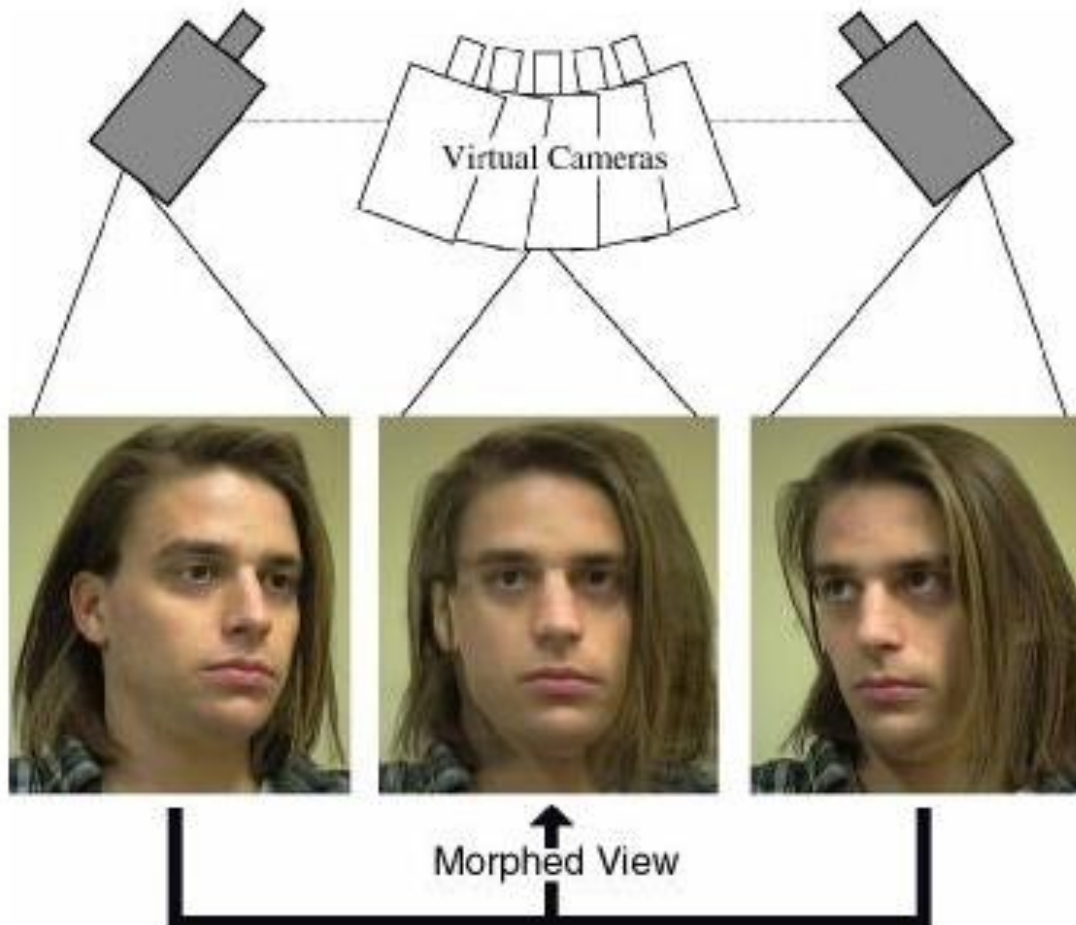


<http://www.cs.cmu.edu/afs/cs/project/VirtualizedR/www/VirtualizedR.html>

[Takeo Kanade *et al.*, CMU]

III. Dispozitive de recunoaștere a gesturilor

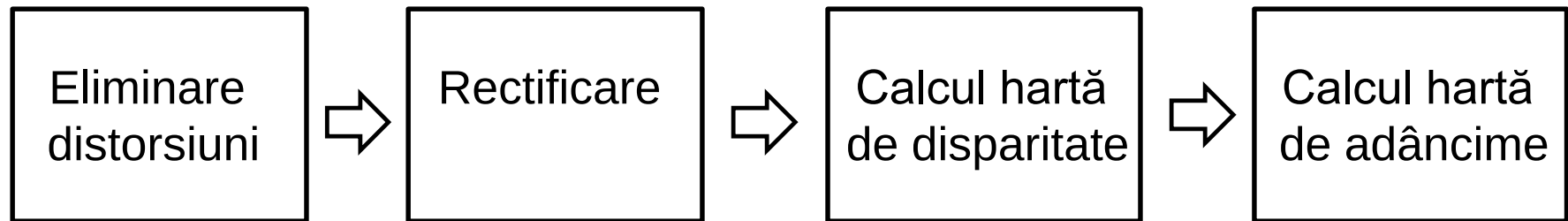
Camere stereo



III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calcul distanță obiecte



III. Dispozitive de recunoaștere a gesturilor

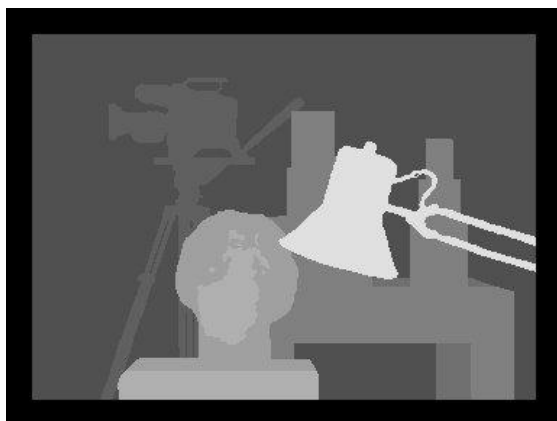
Camere stereo



Cameră
stângă



Cameră
dreaptă



Hartă de adâncime

Cod sursă:

http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_tutorials.html

Baze de date de imagini stereo:

<http://vision.middlebury.edu/stereo/data/>

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calculul distanță obiecte – Eliminarea distorsiuni

Camera
stângă



Camera
dreaptă



III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calculul distanță obiecte - Rectificare

Rectificarea este transformarea prin care o pereche de linii epipolare devin coliniare și paralele pe axa orizontală;

Cameră
stângă



Cameră
dreaptă



[Loop și Zhang,
CVPR'99]

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calculul distanță obiecte - Rectificare

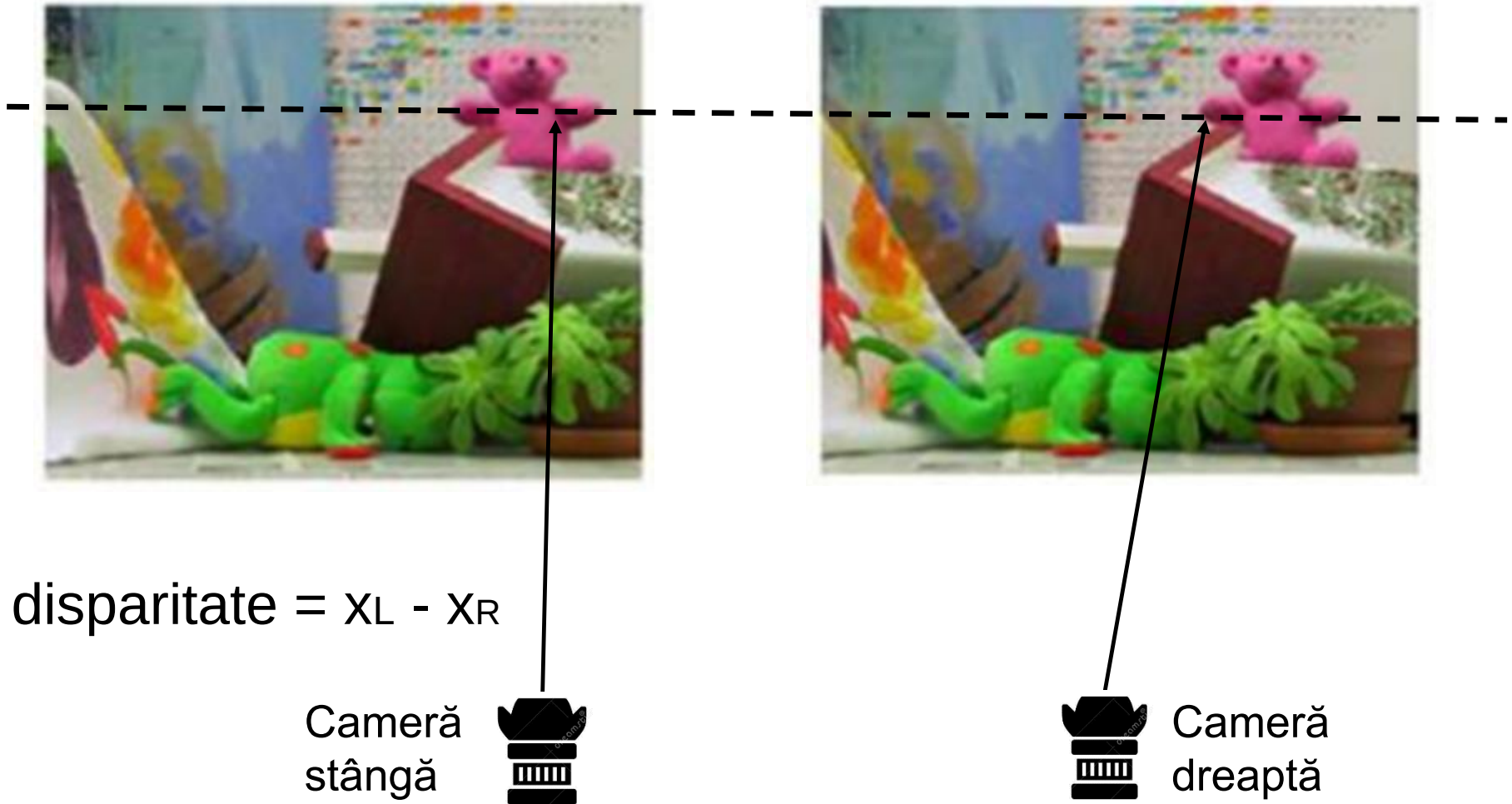


Linii epipolare

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calculul distanță obiecte - Calcul hartă de disparitate



III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calcul distanță obiecte - Calcul hartă de disparitate

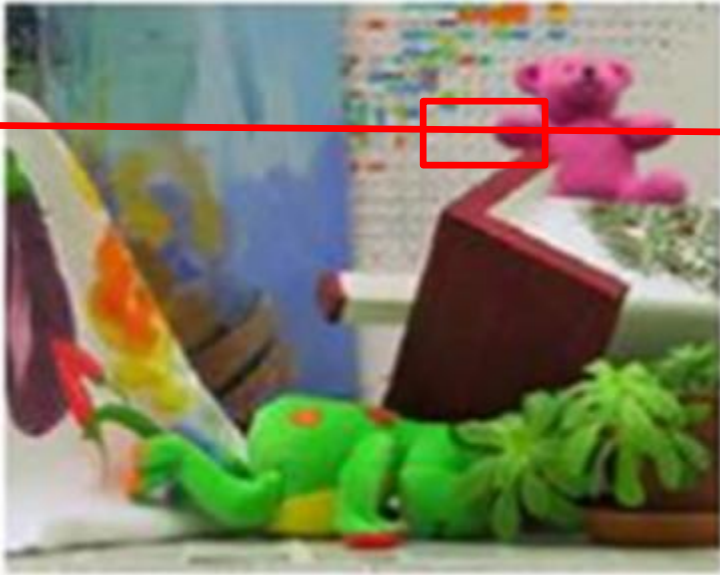
Schemă algoritm

- pentru fiecare linie epipolară
 - pentru fiecare pixel din imaginea dreaptă
 - compară cu fiecare pixel din imaginea dreaptă de pe linia epipolară corespondentă
 - găsește pixelul cu cel mai mic cost de potrivire

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calcul distanță obiecte - Calcul hartă de disparitate



Ferestre de potrivire (utilizând SSD ca și mărime de minimizare a costului - SSD (Sum Squared Differences))

$$E(x, y; d) = \sum_{(x', y') \in N(x, y)} [I_L(x' + d, y') - I_R(x', y')]^2$$

d = disparitate

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

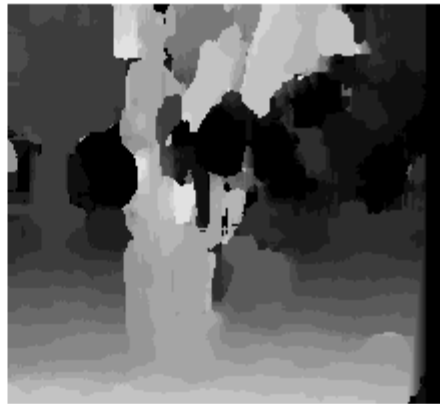
Cât de mare trebuie să fie dimensiunea ferestrei?

Dimensiune mică: multe detalii

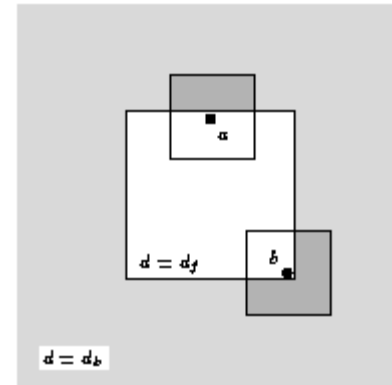
Dimensiune mare: mai puține erori / detalii mai puține



$w = 3$



$w = 20$



III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Alte metode

Raw pixel values (correlation)

Band-pass filtered images [Jones & Malik 92]

“Corner” like features [Zhang, ...]

Edges [many people...]

Gradients [Seitz 89; Scharstein 94]

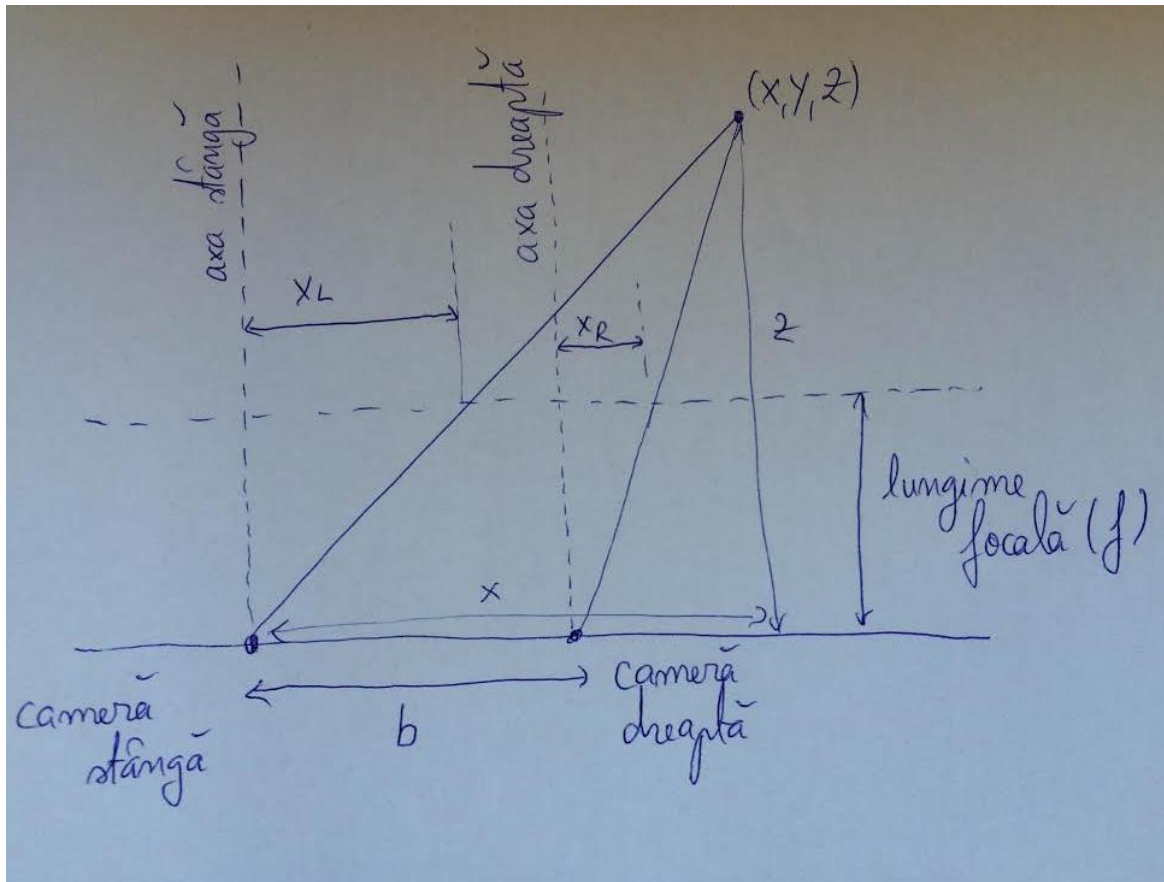
Rank statistics [Zabih & Woodfill 94]

<https://www.eecis.udel.edu/~grauerg/globalStereoPresCompVisionClass.pdf>

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Calcul distanță obiecte - Calcul hartă de adâncime



$$\left. \begin{aligned} \frac{x}{z} &= \frac{x_L}{f} \\ \frac{x - b}{z} &= \frac{x_R}{f} \end{aligned} \right\}$$

$$z = \frac{bf}{x_L - x_R} = \frac{bf}{\text{disparitate}}$$

III. Dispozitive de recunoaștere a gesturilor

Camere stereo

Dezavantaje

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

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

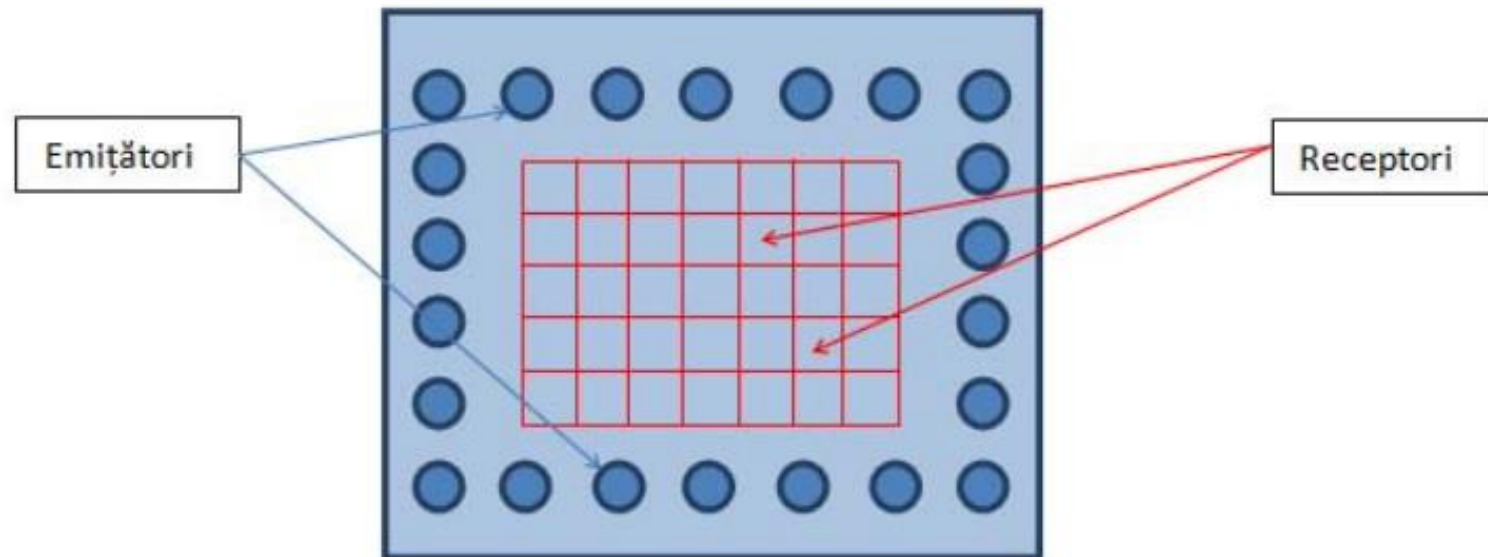
Principiu funcționare

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

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

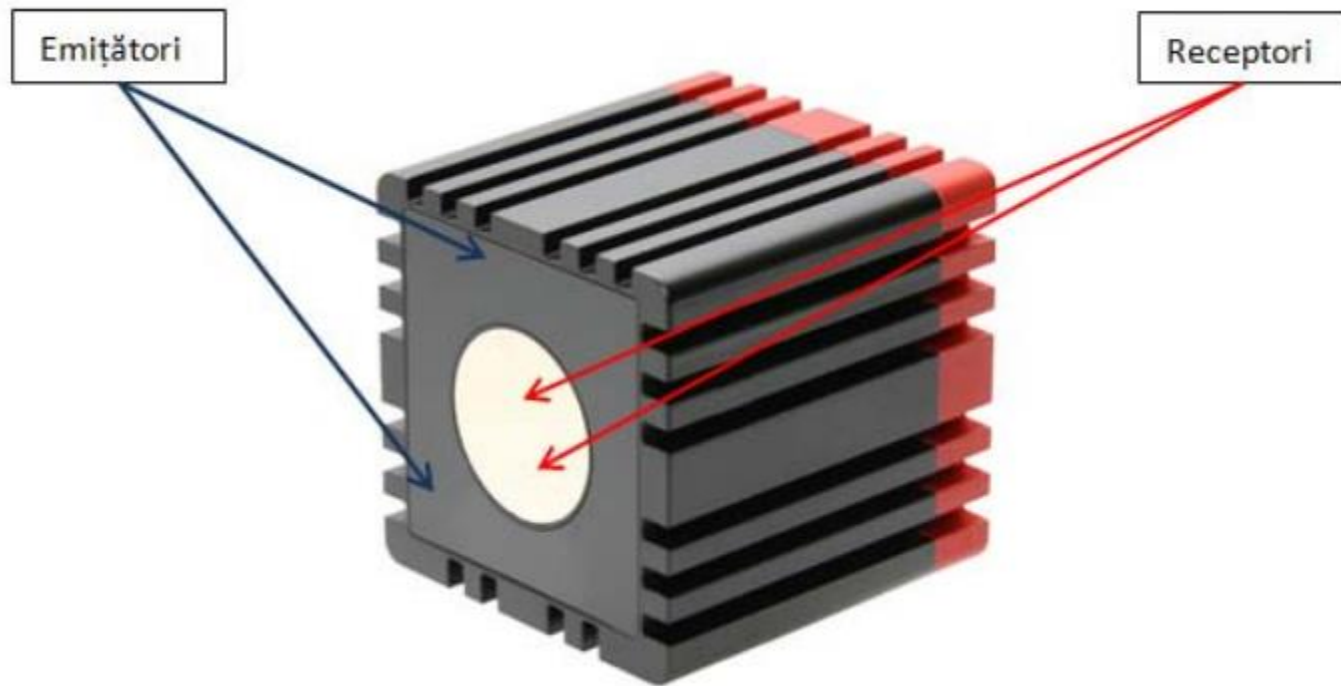
Schema unui senzor ToF matricial



III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

Schema unui senzor ToF matricial



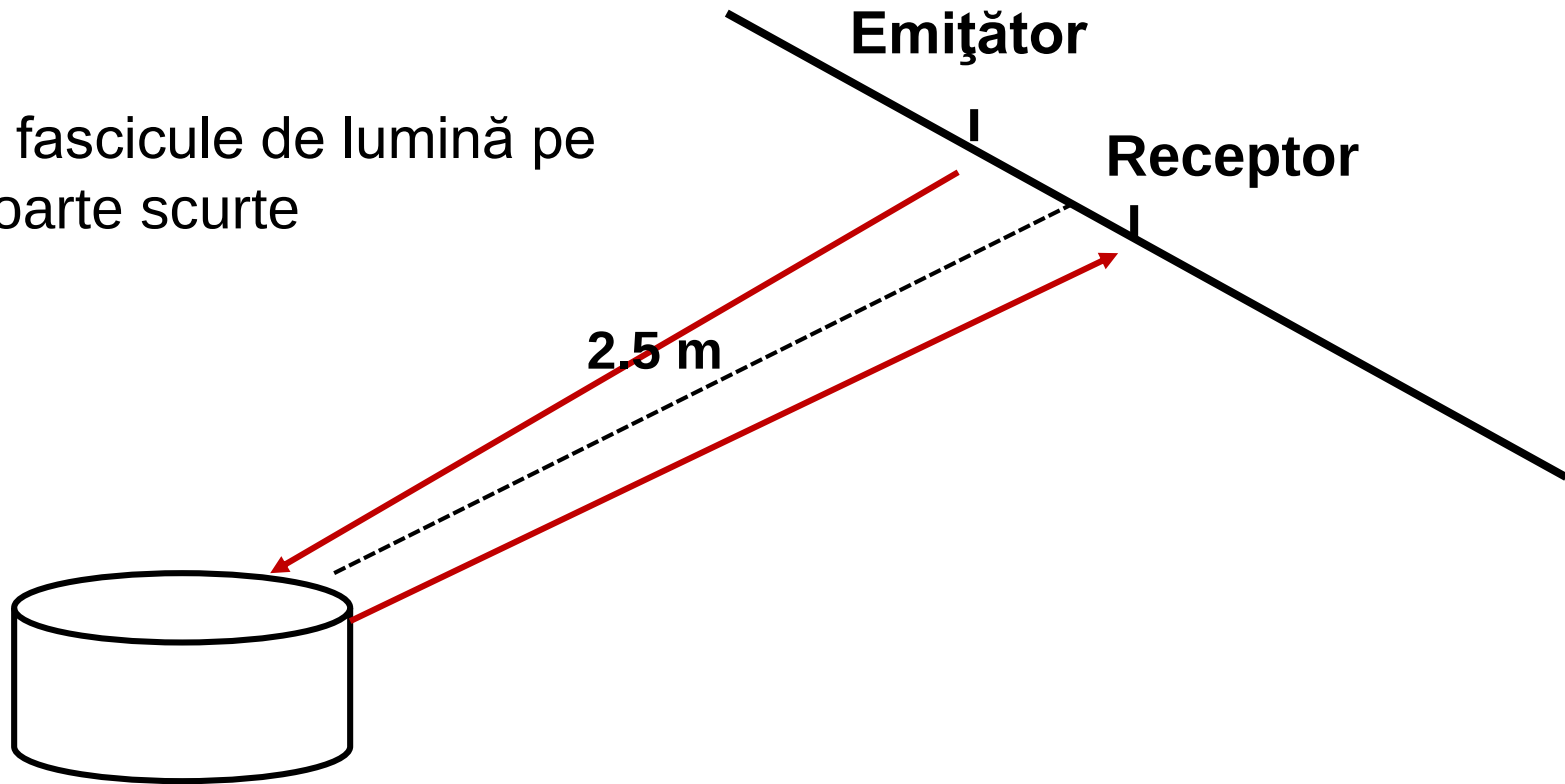
Plasarea emițătorilor și a receptorilor pentru o cameră ToF MESA Imaging 3D SR4000

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

Principiu funcționare

- Se emit fascicule de lumină pe perioade foarte scurte



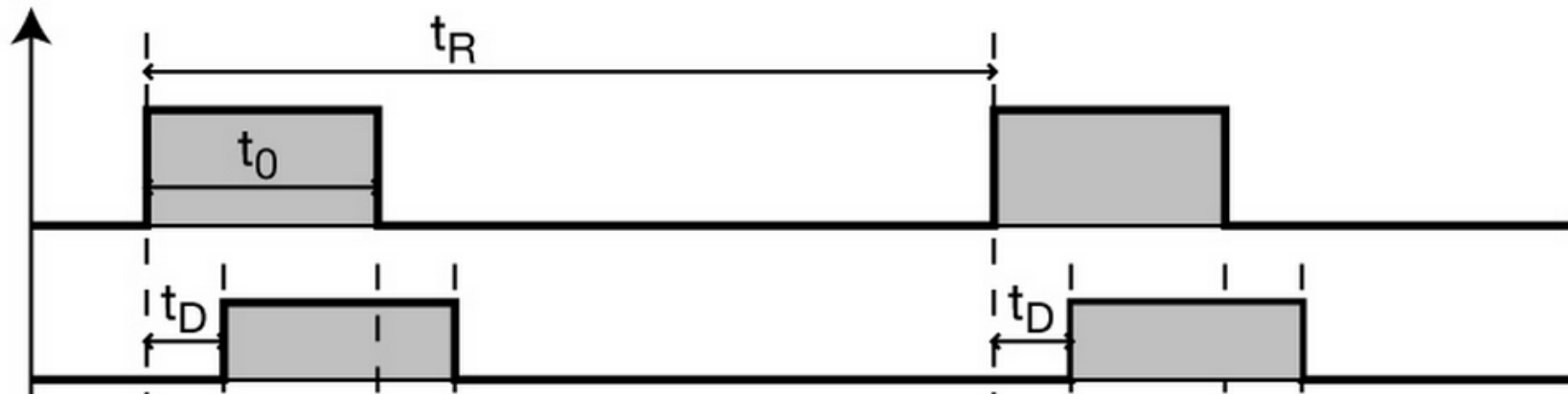
III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

Principiu funcționare

- Având în vedere că viteza luminii este aprox. 300.000.000 m/s, timpul de răspuns se calculează

$$t_D = 2 \cdot \frac{D}{c} = 2 \cdot \frac{2.5 \text{ m}}{300\,000\,000 \frac{\text{m}}{\text{s}}} = 0.000\,000\,016\,66 \text{ s} = 16.66 \text{ ns}$$



III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

Avantaje

➤ **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**

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime TOF

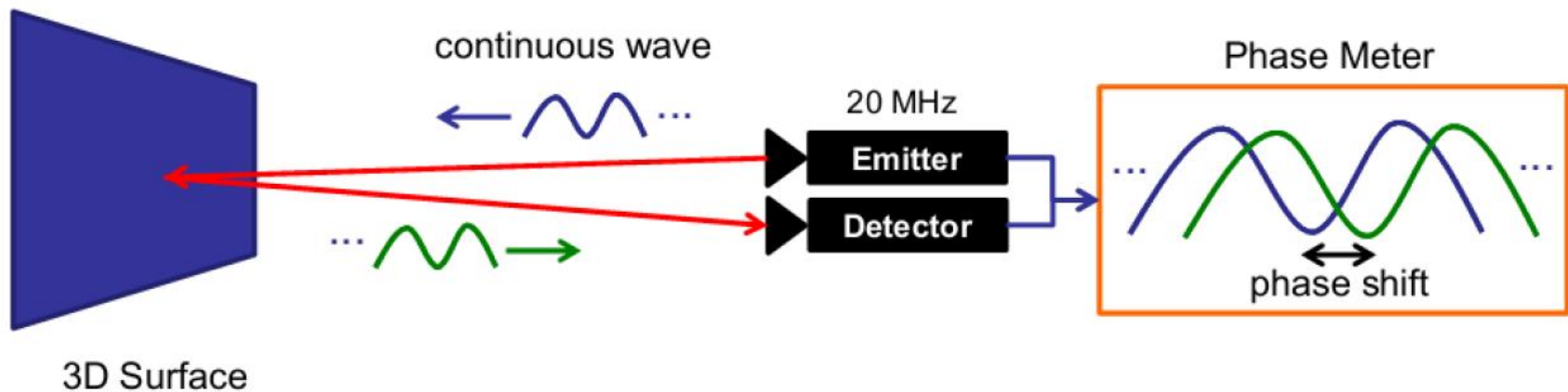
Dezavantaje

- **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.

III. Dispozitive de recunoaștere a gesturilor

Camere de adâncime 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.



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu modulație continuă

- ✓ Semnal sinusoidal emis: $g(t) = \cos(\omega t)$

Semnalul este reflectat de către obiectul din scenă și datorită faptului că o parte din energie este absorbită apare un decalaj.

- ✓ Semnal primit: $s(t) = b + a \cos(\omega t + \phi)$
- ✓ Funcția de corelație dintre cele două semnale (TTI):

$$c(\tau) = s * g = \int_{-\infty}^{\infty} s(t) \cdot g(t + \tau) dt$$

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu modulație continuă

- ✓ Funcția de corelație dintre cele două semnale (TTI):

$$c(\tau) = s * g = \int_{-\infty}^{\infty} s(t) \cdot g(t + \tau) dt$$

- ✓ Care devine pentru cazul nostru:

$$c(\tau) = \frac{a}{2} \cos(\omega\tau + \phi) + b$$

ω – frecvența
 a - amplitudinea

- ✓ Se calculează mai multe valori ale lui $c(\tau)$

$$A_i = c(i \cdot \pi/2), \quad i = 0, \dots, 3$$

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu modulație continuă

✓ În final obținem:

$$\phi = \arctan2(A_3 - A_1, A_0 - A_2)$$

$$a = 1/2 \sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}$$

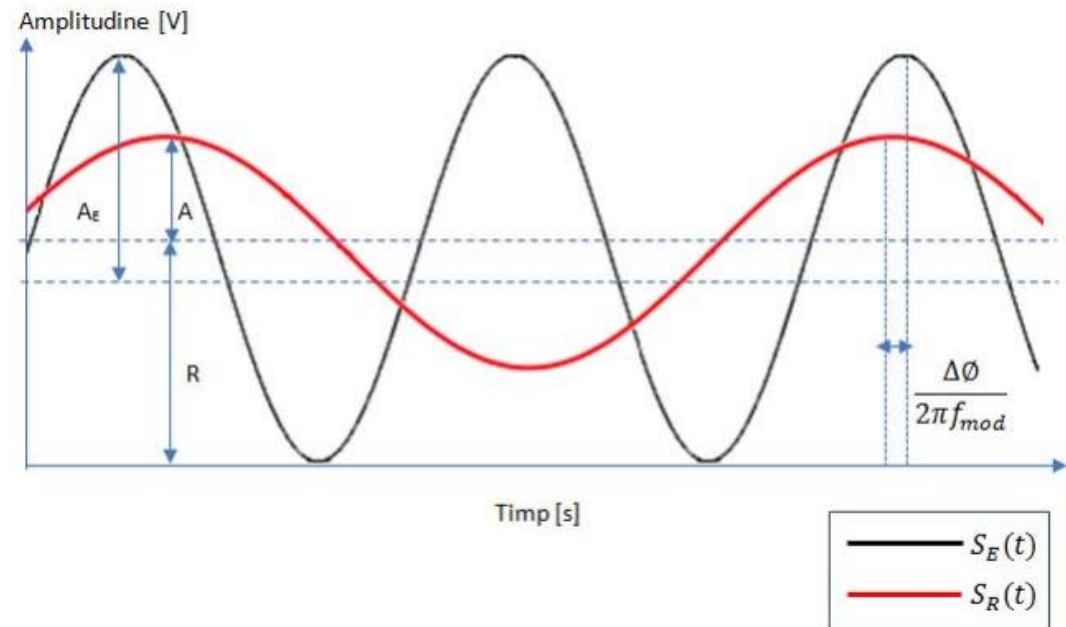
$$d = \frac{c}{4\pi\omega} \phi$$

b: constant bias

a: amplitude

ϕ : phase shift

τ : internal offset



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu modulație continuă

➤ Avantaje

- ✓ Pot exista mai multe surse de lumină simultane;
- ✓ Se pot aplica diferite metode de modulație.

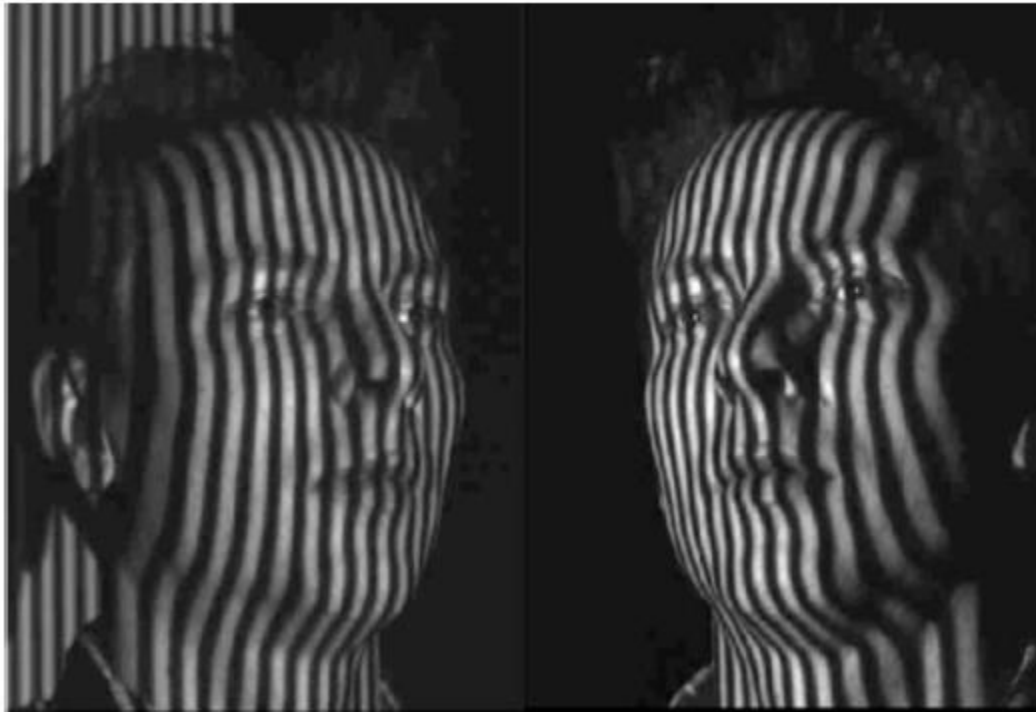
➤ Dezavantaje

- ✓ Sensibile la zgomot;
- ✓ Viteză mai scăzută;
- ✓ Semnale perfect sinusoidale sunt greu de obținut în practică.

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

- ✓ Proiectează o textură luminoasă care este preluat de receptor.
- ✓ Distorsiunile apărute ajută la calculul structurii 3D.



III. Dispozitive de recunoaștere a gesturilor

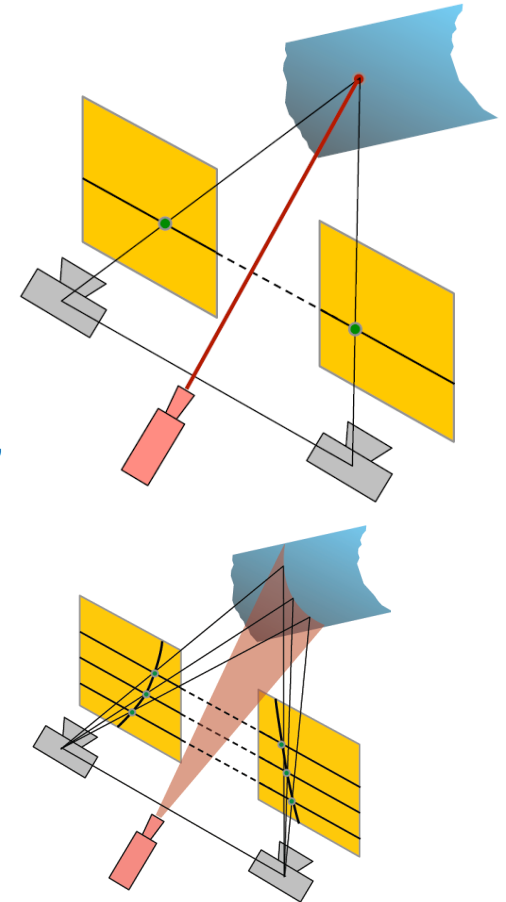
Camere TOF cu lumină structurată

- Marea problemă pentru camerele stereo este nevoia de corespondență dintre diversele poziții din imagini;
- Căutarea corespondenței este un lucru dificil (texturi / culori mate);
- Ideea: simplificare algoritmului de corespondență prin utilizarea iluminării;
- Principiu de bază: utilizarea proiectorului pentru a crea corespondențe.

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

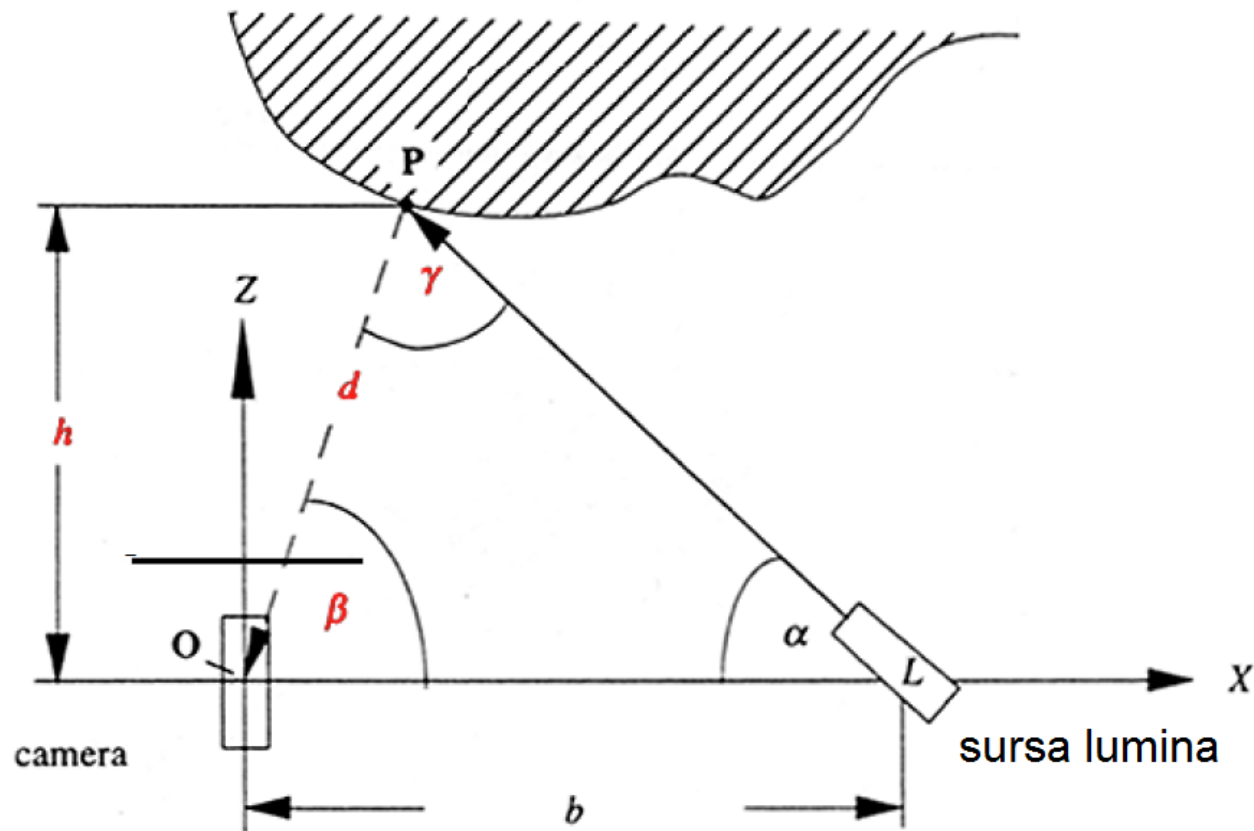
- Dacă proiectăm un singur punct atunci realizarea corespondenței este simplă;
- Prin analogie, dacă proiectăm o dreaptă se poate realiza corespondența



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

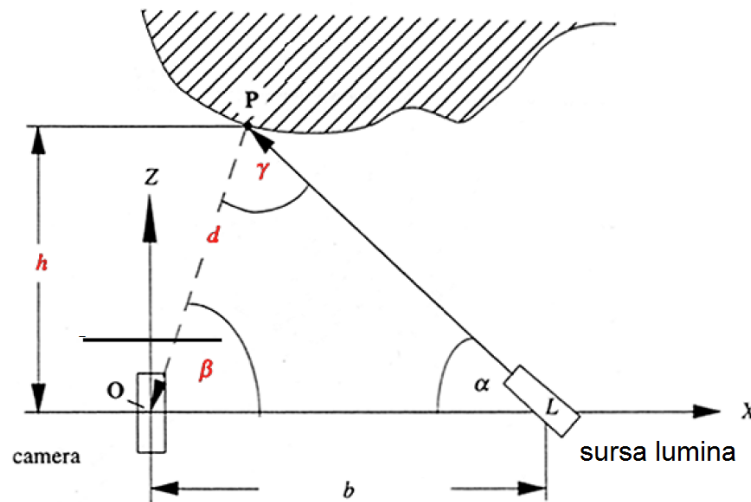
- Distanța pentru un singur punct



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

➤ Distanța pentru un singur punct (2D)



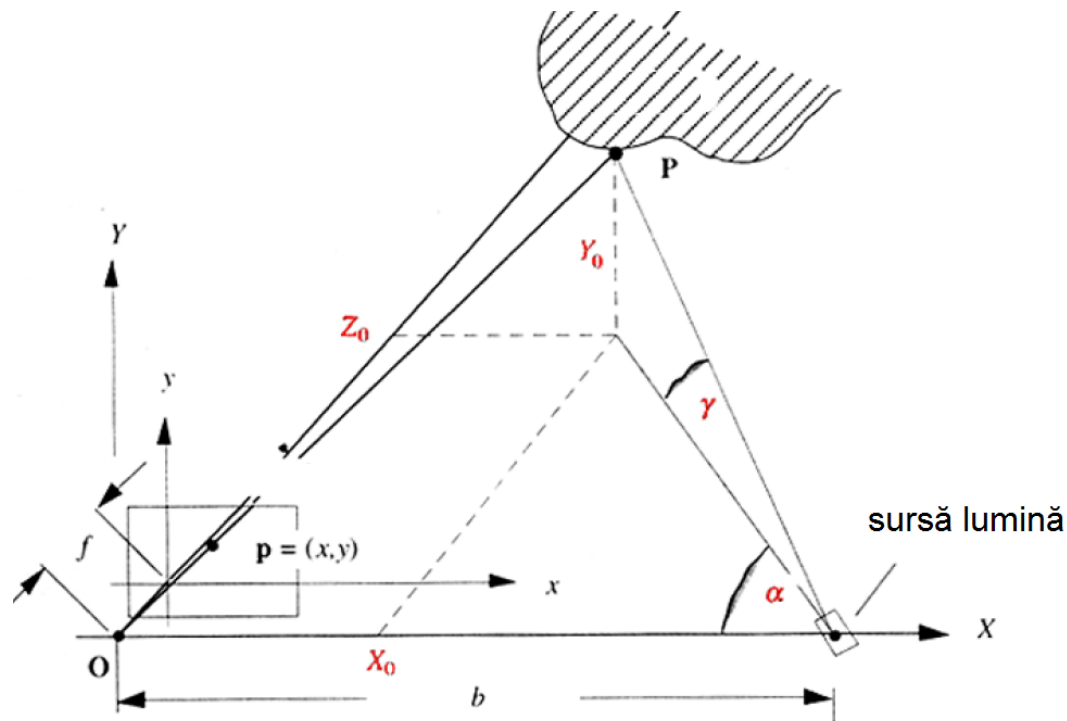
Punctele O, L, P realizează un triunghi. Utilizând teorema sinusurilor vom avea:

$$\frac{d}{\sin \alpha} = \frac{b}{\sin \gamma} \quad \longrightarrow \quad d = \frac{b \cdot \sin \alpha}{\sin \gamma} = \frac{b \cdot \alpha}{\sin(\pi - \alpha - \beta)} = \frac{b \cdot \alpha}{\sin(\alpha + \beta)}$$

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

- Distanța pentru un singur punct (3D)

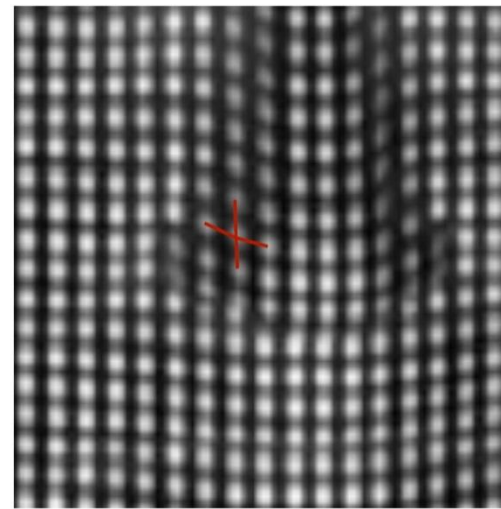
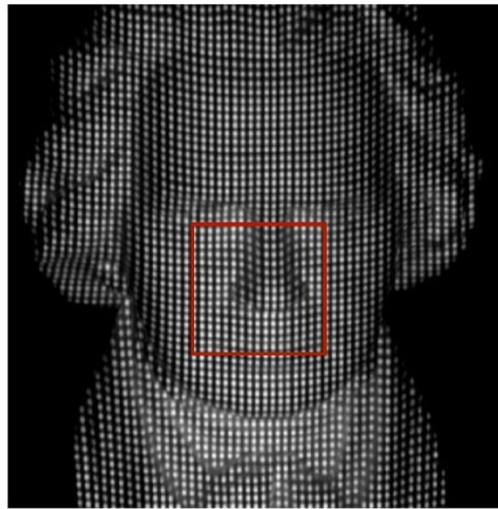


$$X = \frac{\tan \alpha \cdot b \cdot x}{f + x \cdot \tan \alpha}, \quad Y = \frac{\tan \alpha \cdot b \cdot y}{f + x \cdot \tan \alpha}, \quad Z = \frac{\tan \alpha \cdot b \cdot f}{f + x \cdot \tan \alpha}$$

III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

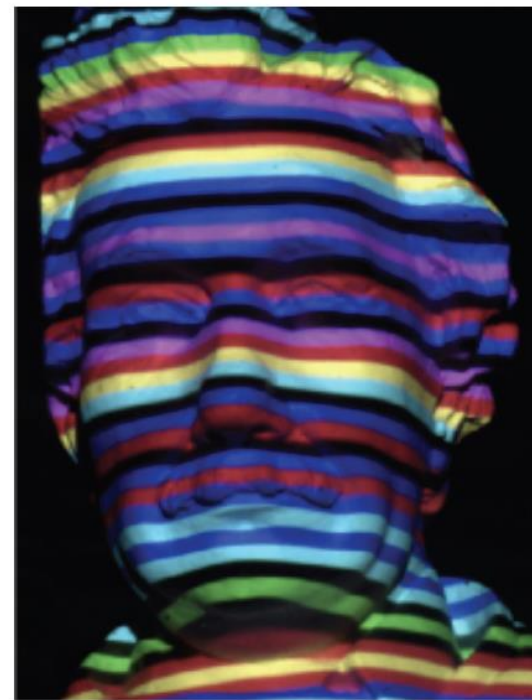
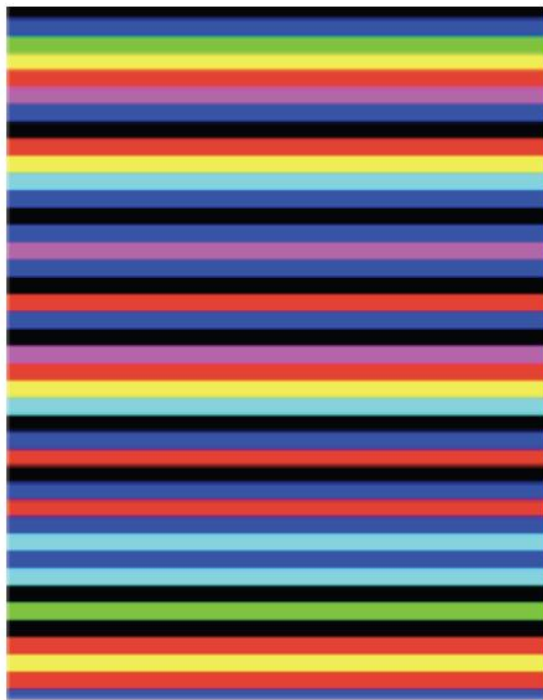
- Perechi de linii
- Grid-uri



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

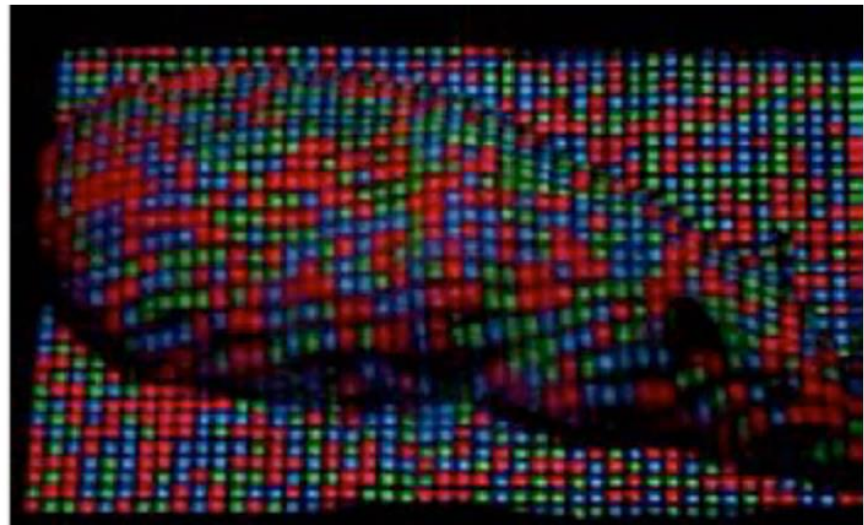
- Linii colorate (coduri De Bruijn)



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

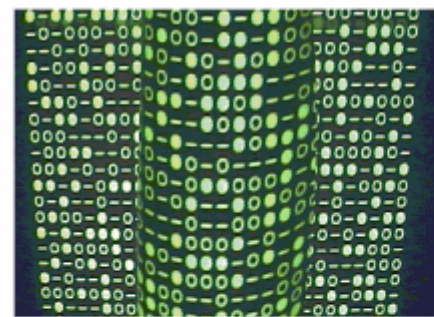
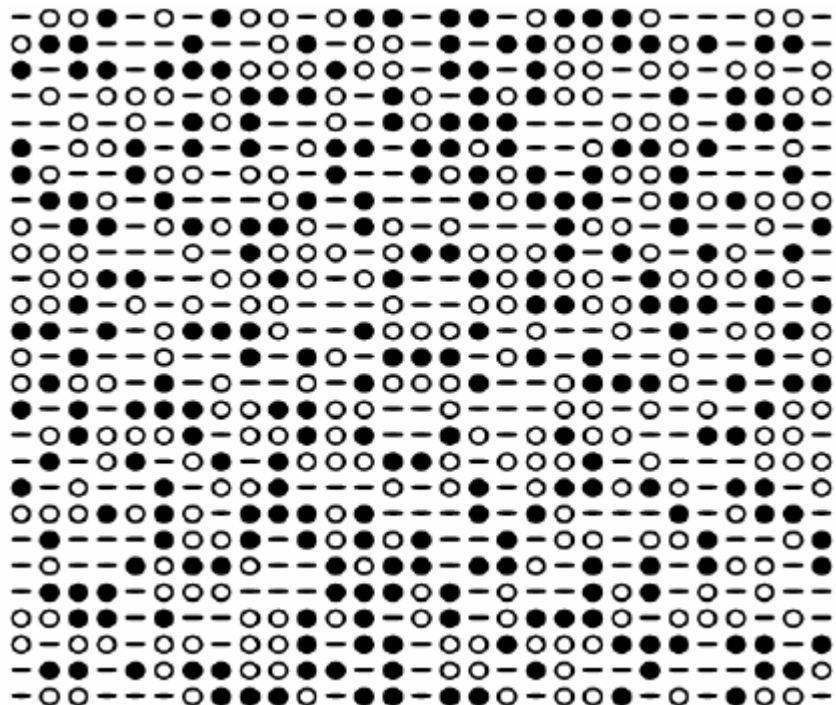
- Puncte colorate



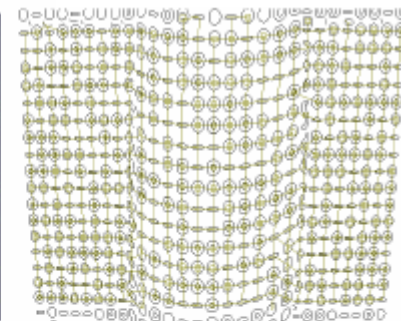
III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

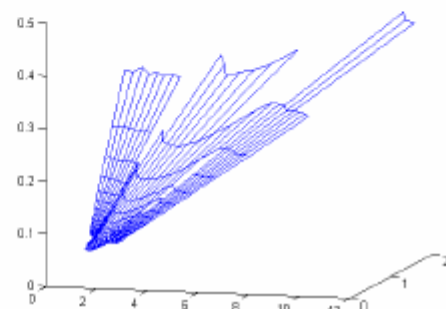
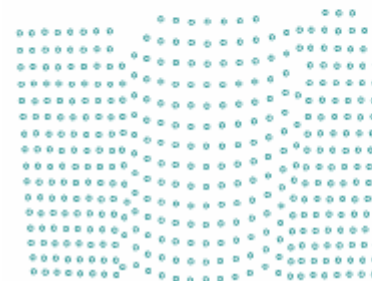
➤ Pattern-uri de puncte



(a)



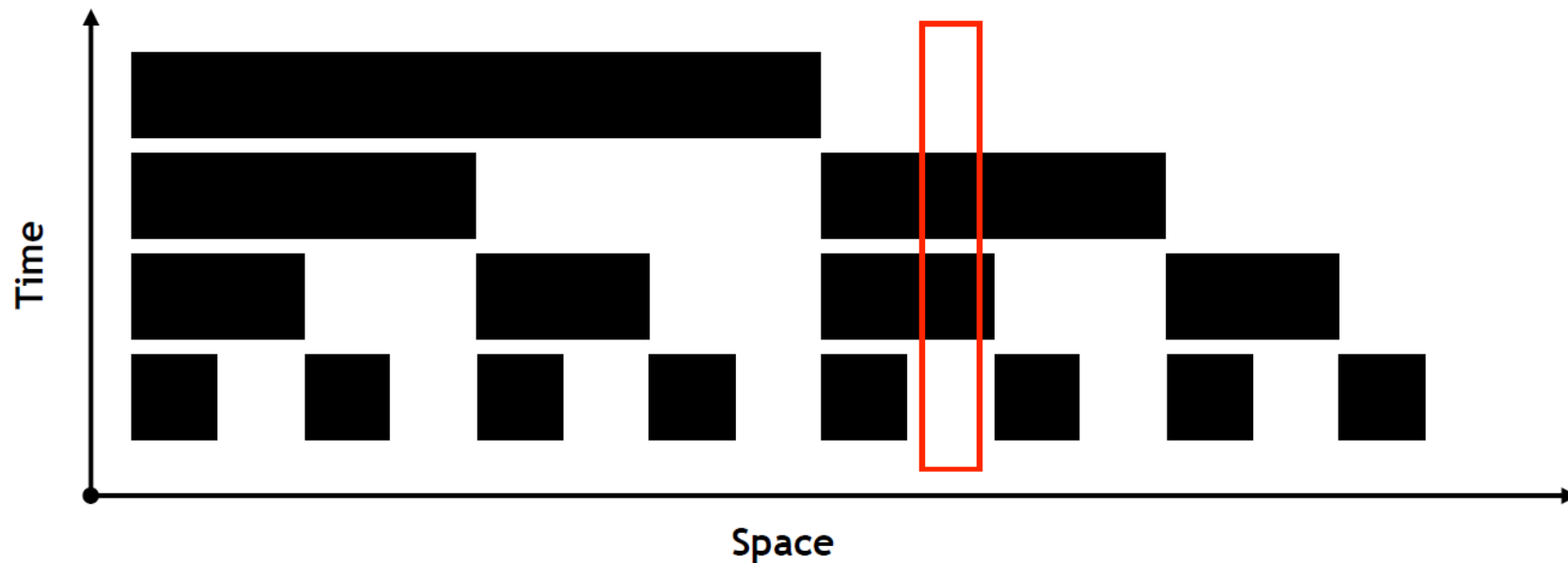
(b)



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

- Multiplexare de coduri în timp



III. Dispozitive de recunoaștere a gesturilor

Camere TOF cu lumină structurată

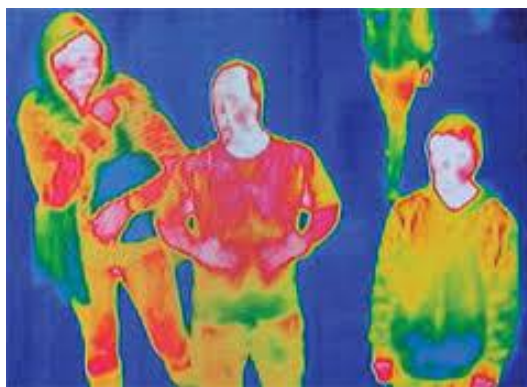
Descriere mai aprofundată în:

Salvi, J., S. Fernandez, et al. (2010). "A state of the art in structured light patterns for surface profilometry." *Pattern Recognition* **43**(8): 2666-2680

III. Dispozitive de recunoaștere a gesturilor

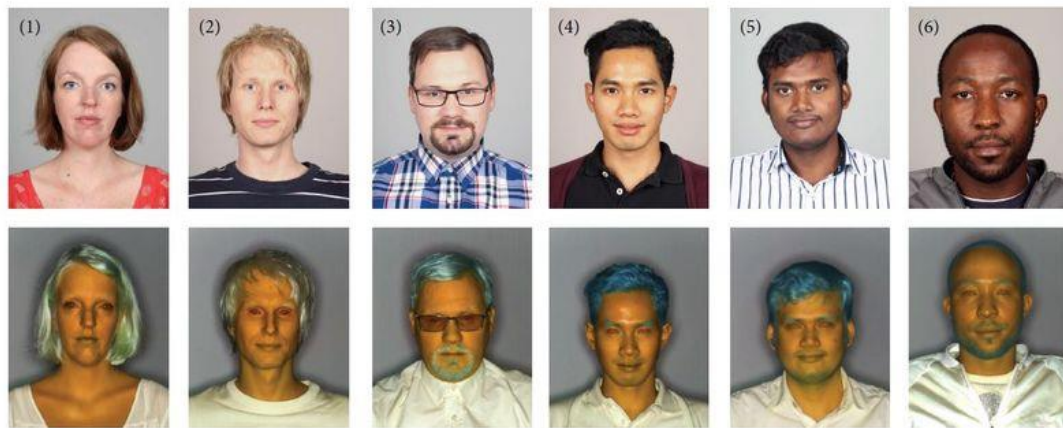
Camere termice

- detectează radiația de infraroșu (temperatura) de la obiectul detectat și îl convertește într-un semnal electronic pentru a genera o imagine termică;
- operează de lungimi de undă de 14,000 nm (față de spectrul vizibil de 450–750 nm).



III. Dispozitive de recunoaștere a gesturilor

Camere termice



III. Dispozitive de recunoaștere a gesturilor

Consola 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-ul;
 - reprezintă o interfață naturală om – mașină;
 - se pot utiliza gesturile și comenzi vocale.



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

- Kinect-ul este un dispozitiv periferic, care le permite utilizatorilor să se folosească de consola Xbox 360 / One fără ajutorul unui controller fizic, ci doar folosind gesturi ale corpului și comenzi vocale;
- A fost inaugurat în data de 01 iunie 2009, sub denumirea de Project Natal;
- Este valabil pentru toate versiunile Xbox360, cele noi având conector special, iar pentru cele vechi cu USB;
- Din februarie 2012 este disponibil și pe PC;
- Din iulie 2014 este disponibilă și versiunea V2 (Kinect One).



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Alte tipuri de console

OpenNI (Asus)



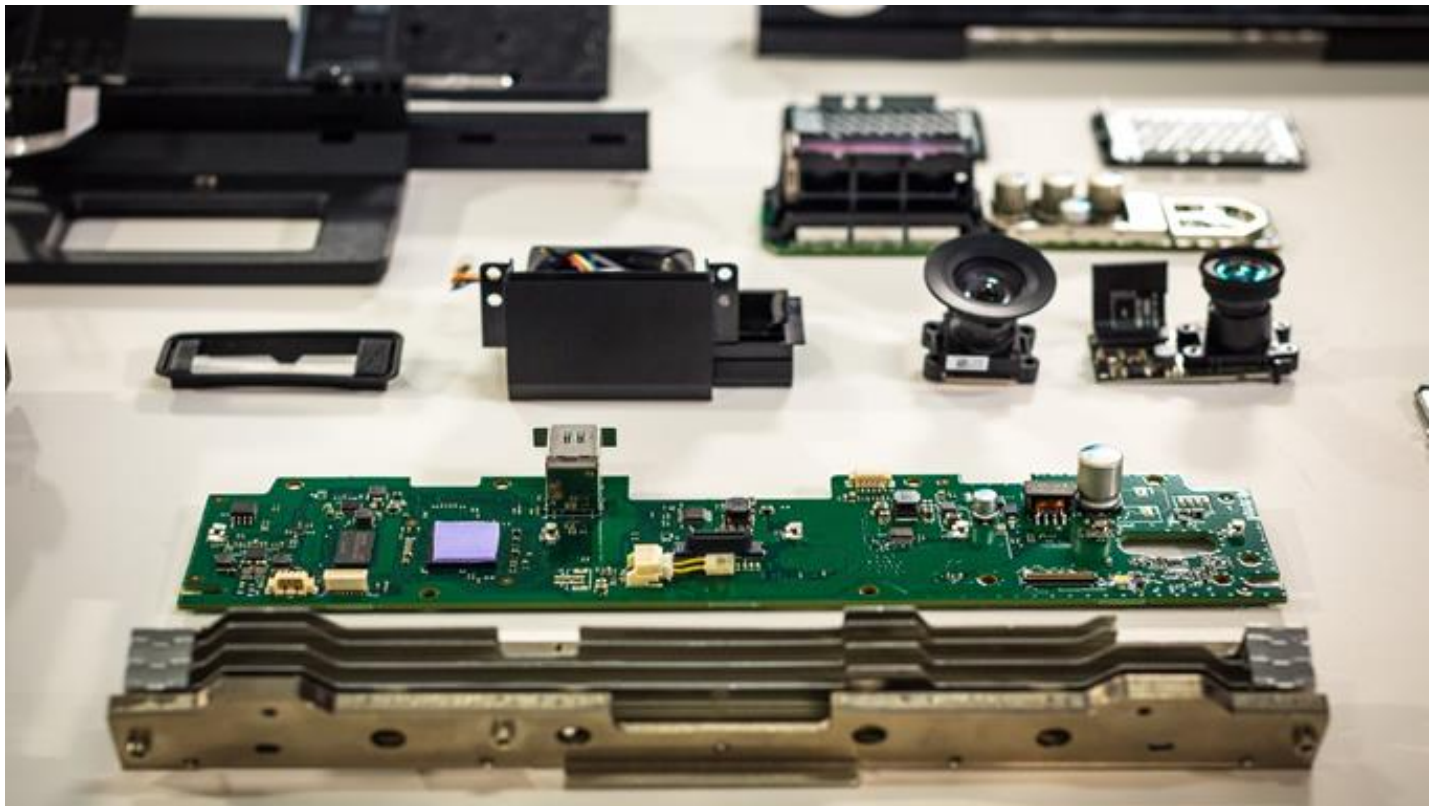
Open Kinect



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Structură Kinect



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Emitor IR

Camera RGB

se poate antrena pentru
detectie de fețe

Senzor de adâncime (receptor IR)

- măsoară distanța față de obiectele din cadru
- captează informația de mișcare

Microfon Multi-array

- permite recunoașterea și
urmărirea sursei audio /
recunoaștere de voce

Power Light



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Ce își propune să rezolve un Kinect?

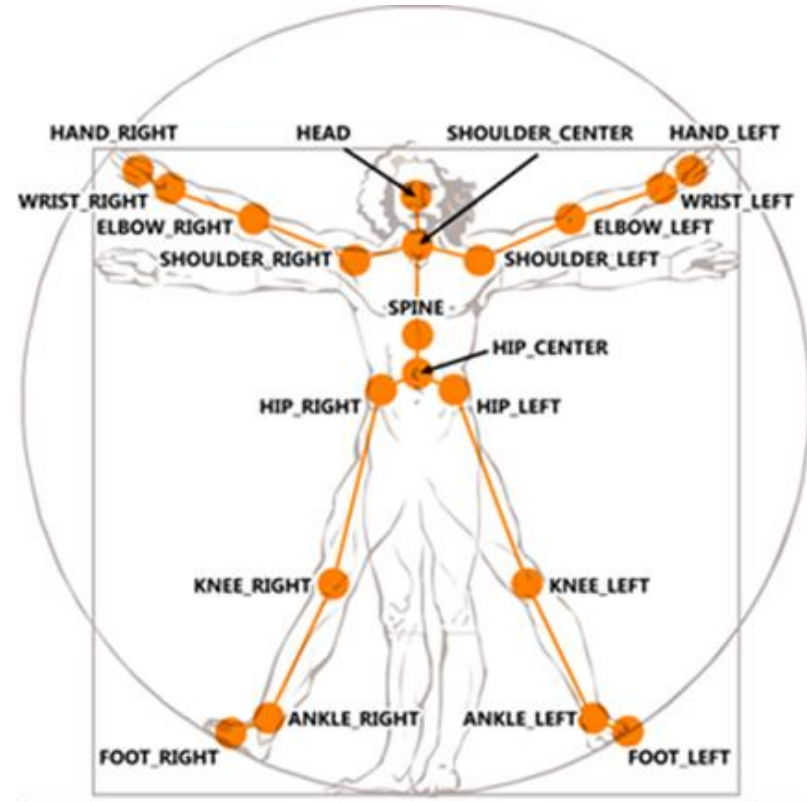
- detecție persoane în scenă ignorând informația de fundal;
- detecție părți componente ale mâinii;
- detecție și urmărire de gesturi;
- un prim pas către recunoaștere de fețe;
- recunoaștere de comenzi vocale.

III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Recunoaștere skeleton

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

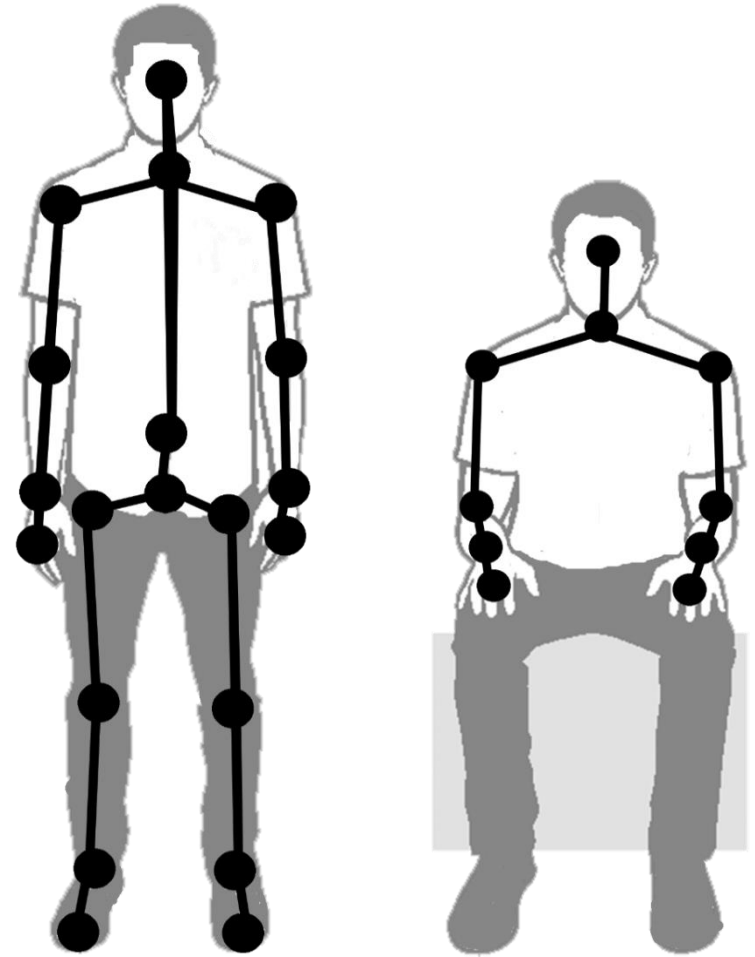


III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Recunoaștere skeleton

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

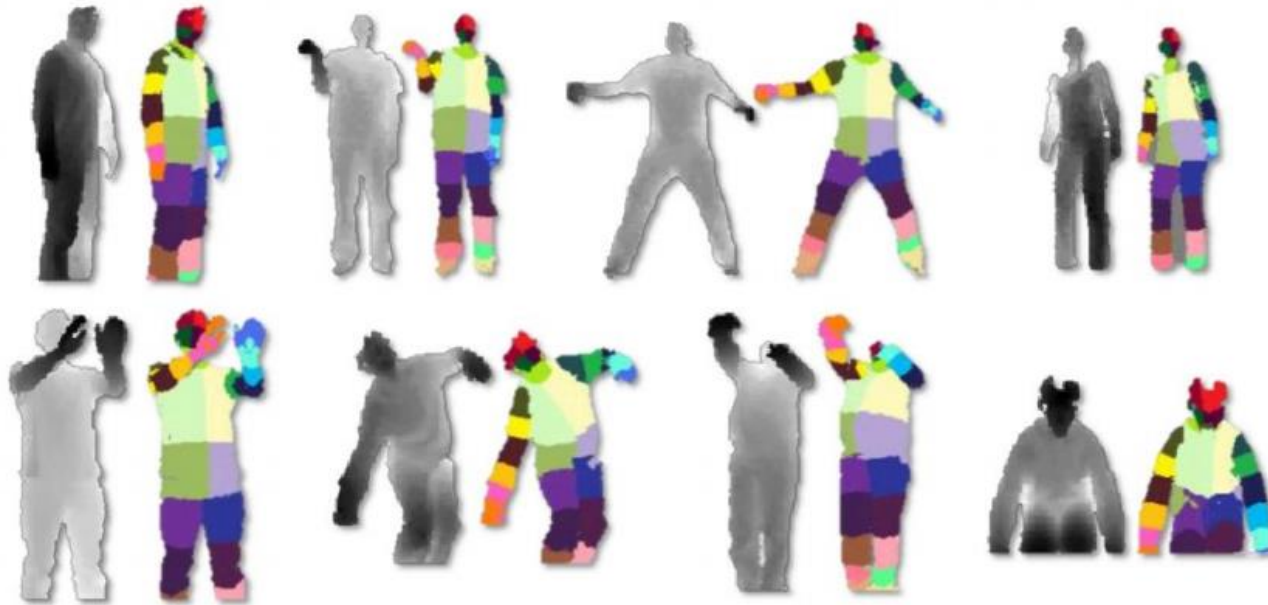


III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Recunoaștere skeleton

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



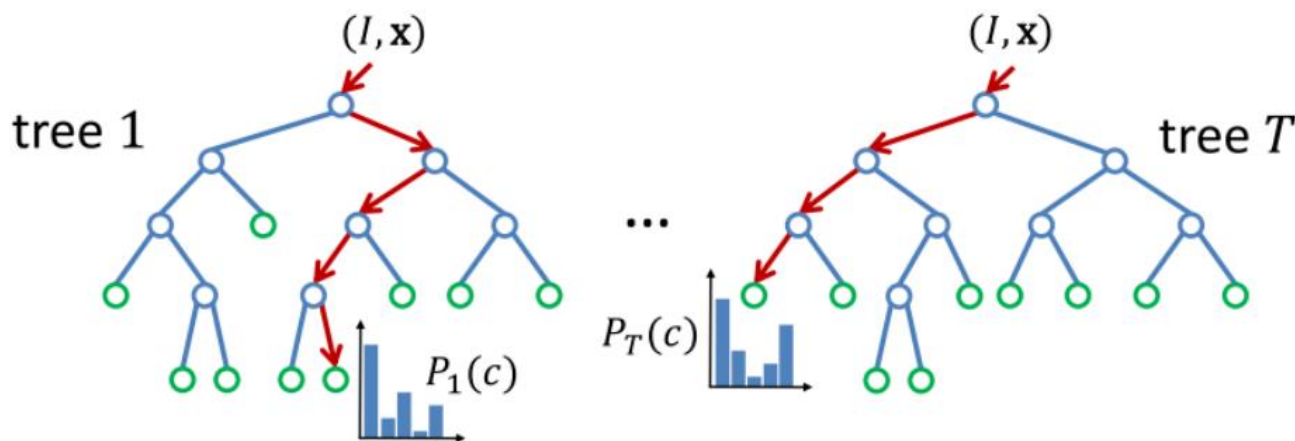
Shotton et al, CVPR(2011)

III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Recunoaștere skeleton

- Algoritmul „random forests” reprezintă o versiune mai sofisticată a algoritmului clasic de arbori de decizie.



Pentru a antrena algoritmul de „random forests” este nevoie de 24.000 ore pe un singur CPU, dar poate dura o zi daca sunt utilizate mii de calculatoare legate în rețea (cloud).

Shotton et al, CVPR(2011)

II. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Comparație Kinect V1 vs Kinect V2

Caracteristică	Kinect V1	Kinect V2
Distanță adâncime	0.4m → 4.0m	0.4m → 4.5m
Rezoluție color	640×480	1920×1080
Rezoluție adâncime	320×240	512×424
Rezoluție infraroșu	None	512×424
Microfoane	4-mic array	4-mic array
USB	2.0	3.0

III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

Comparație Kinect V1 vs Kinect V2

- Kinect V1 poate ține tracking-ul la 40 de puncte pe față
- Kinect V2 poate recunoaște și câteva expresii faciale simple, activități și accesorii:

- ✓ ochi închiși
- ✓ direcția ochilor
- ✓ gură deschisă
- ✓ existență ochelari
- ✓ fericire
- ✓ expresie neutră



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

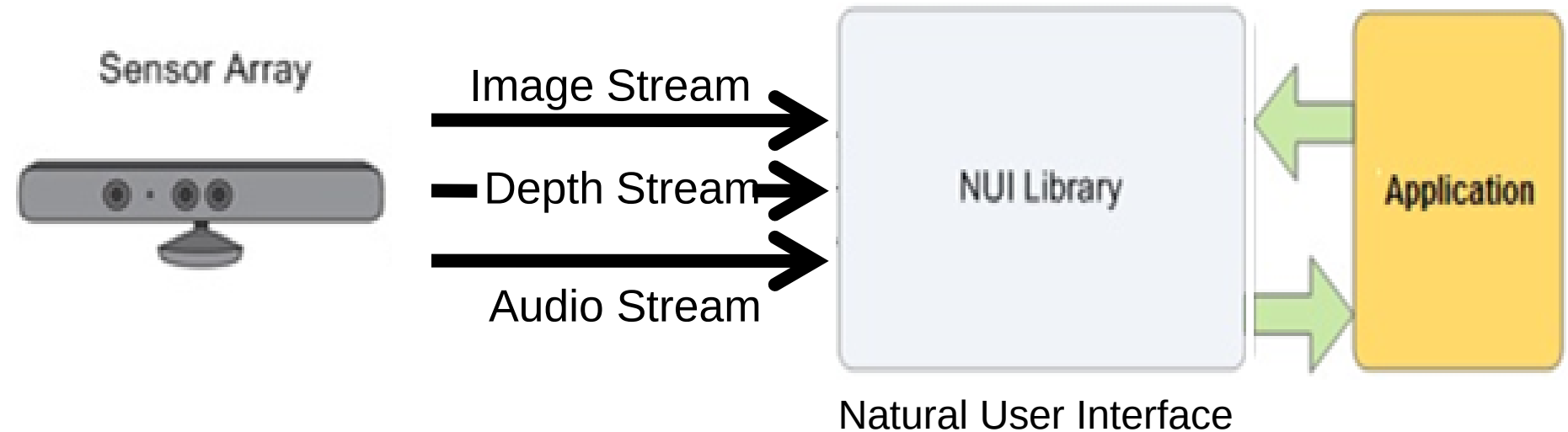
Windows SDK

- Kinect software development kit (SDK) pentru Windows a fost creat inițial pentru Windows 7 în iunie 2011;
- SDK-ul include drivere Windows 7 compatibile PC către dispozitivul Kinect (!nu merge pe Windows XP);
- Permite capabilități de dezvoltare a aplicațiilor cu C++, C# sau Visual Basic cu Microsoft Visual Studio 2010;
- Include clase pentru: captarea semnalelor senzorilor atașați, **skeletal tracking**, capabilități audio și **exemple de cod și documentație**.

III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

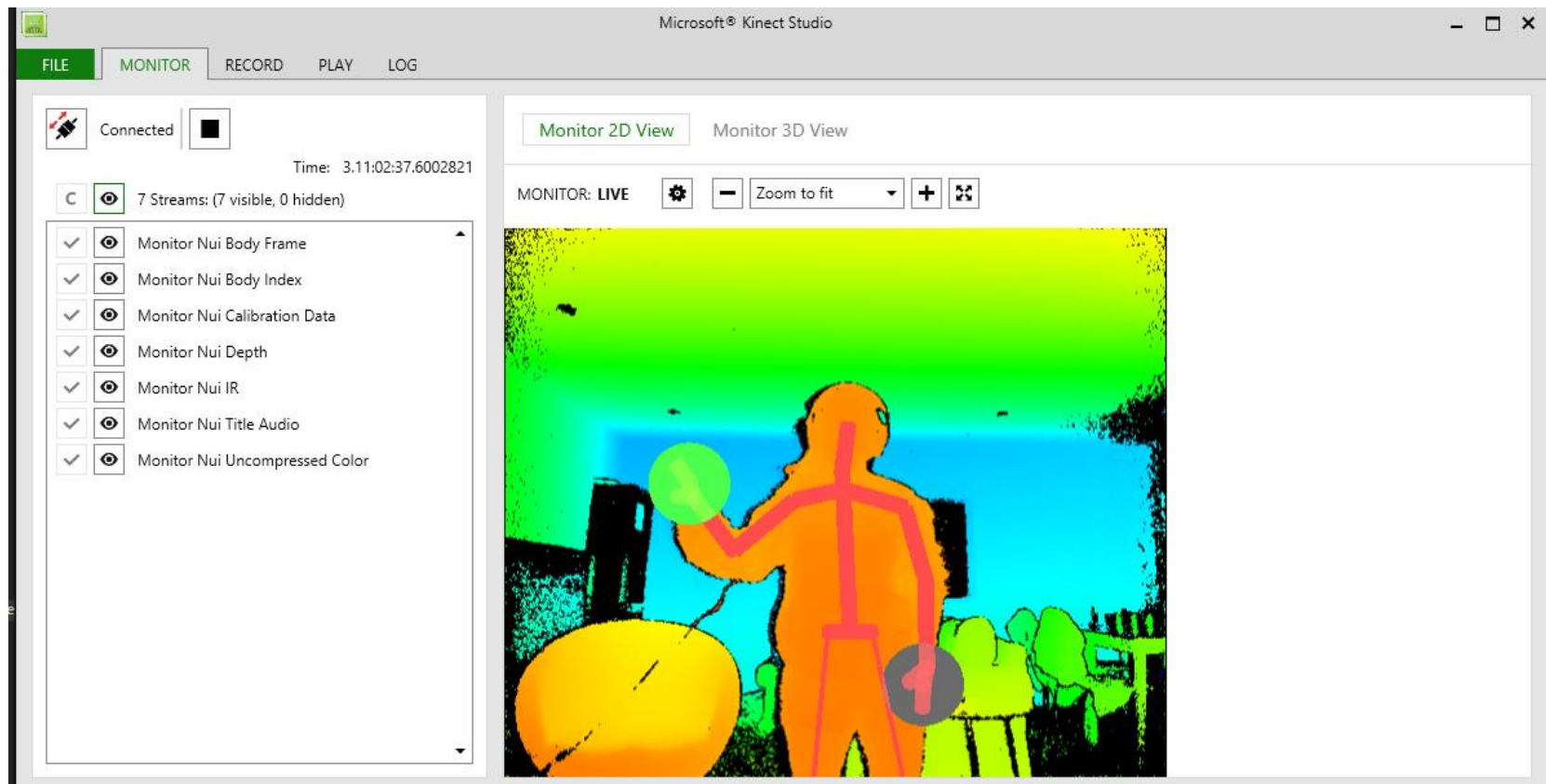
Windows SDK



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

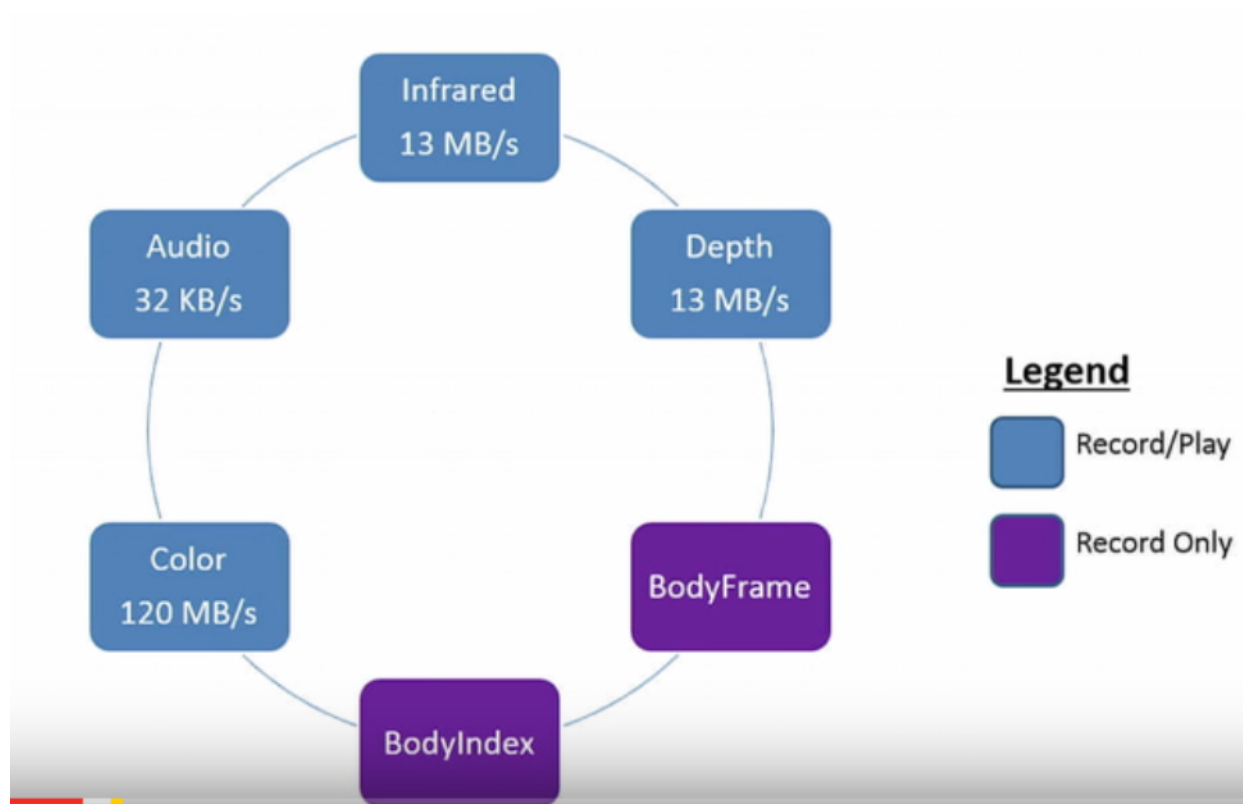
Kinect Studio



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

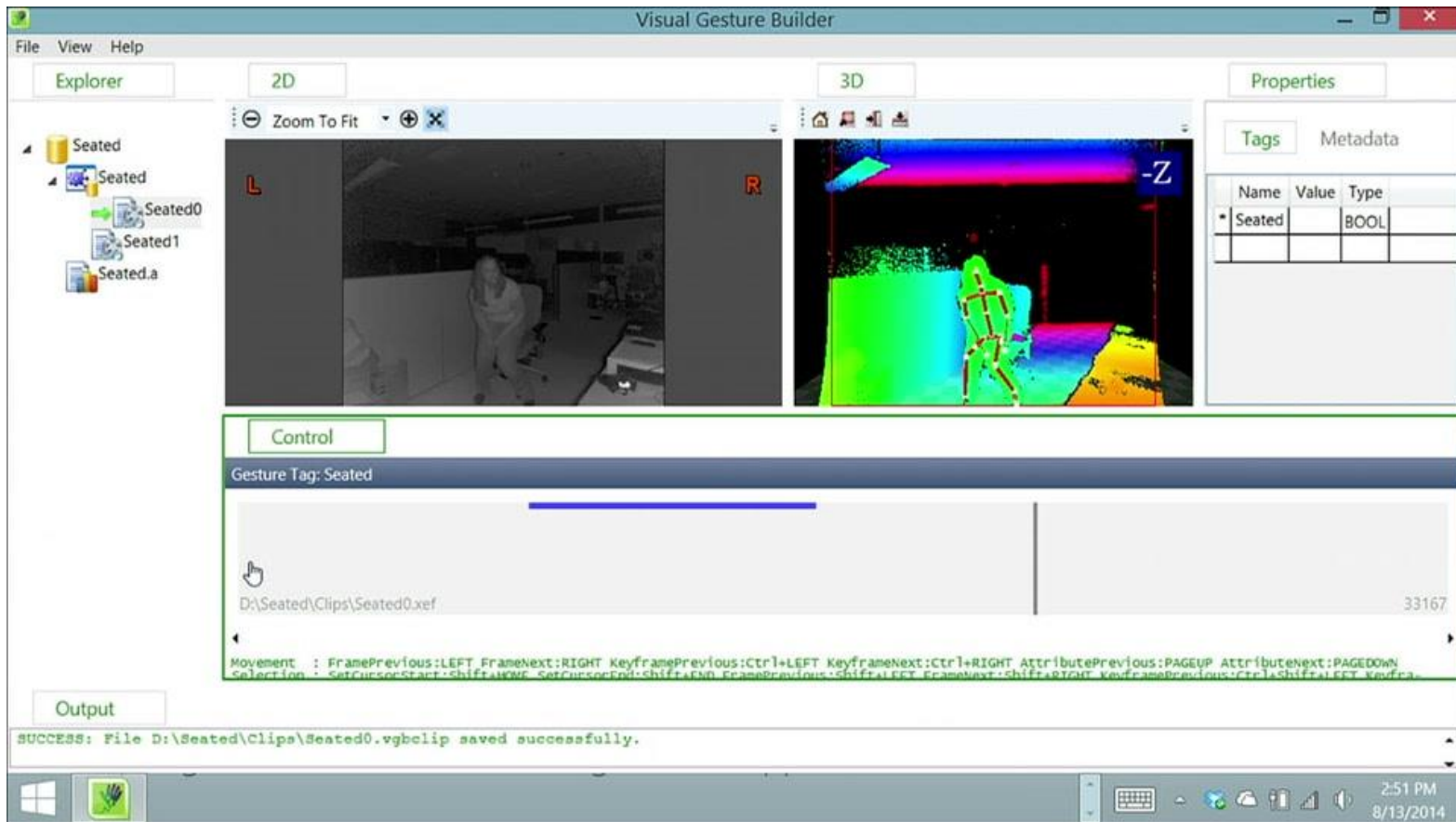
Kinect Studio



III. Dispozitive de recunoaștere a gesturilor

Consola Kinect

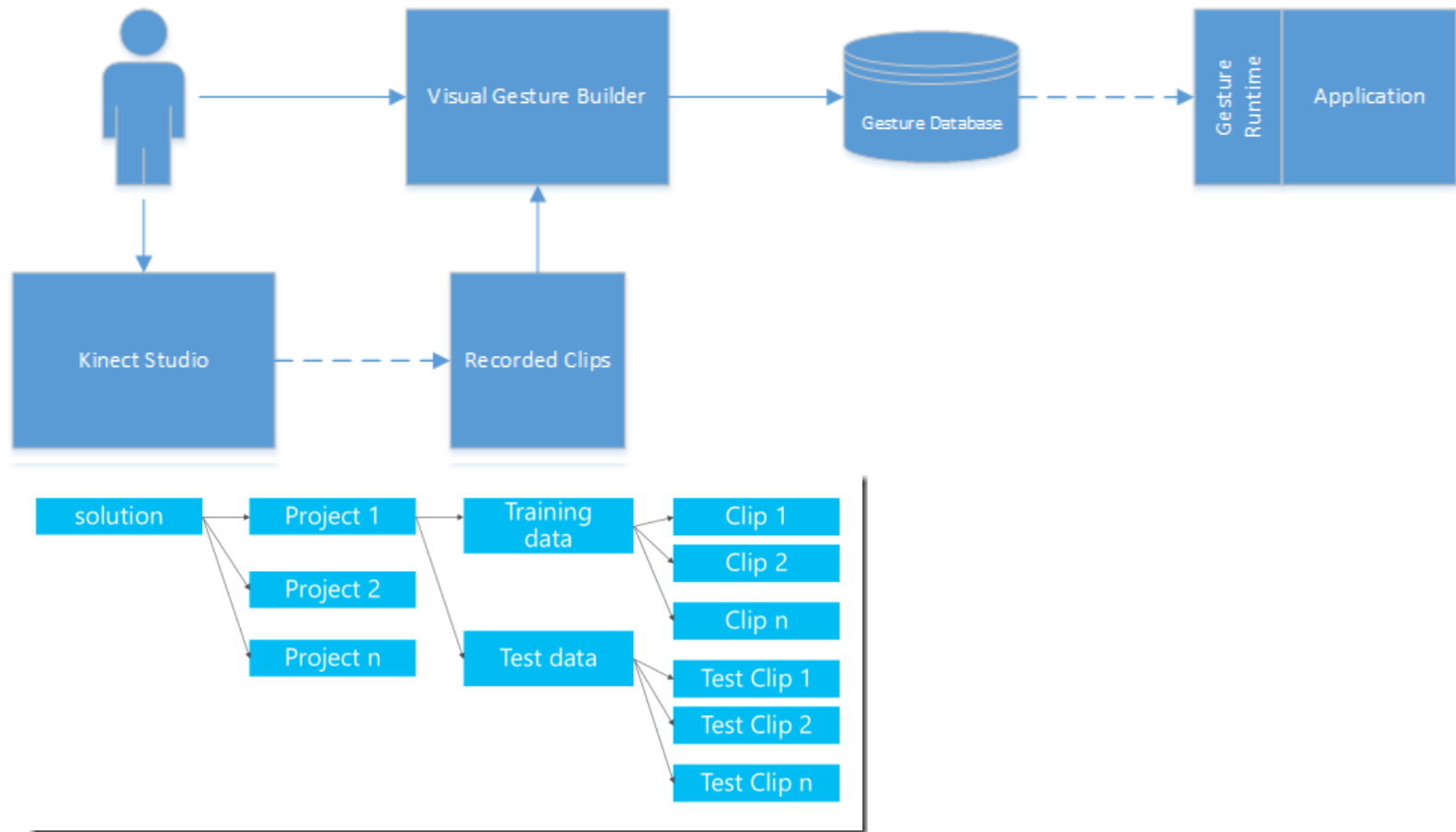
Visual Gesture Builder



III. Dispozitive de recunoaștere a gesturilor

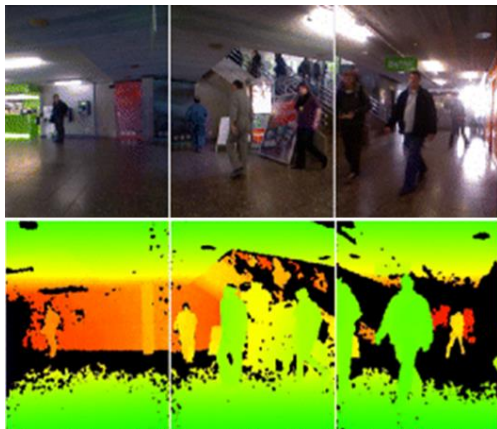
Consola Kinect

Antrenare gesturi



III. Dispozitive de recunoaștere a gesturilor

Baze de date RGBD



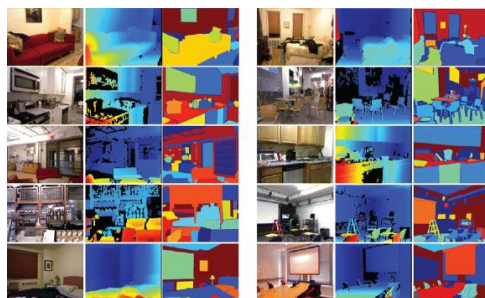
RGB-D people dataset



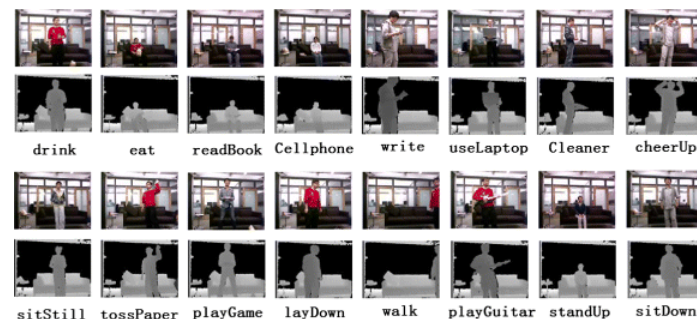
Object recognition



Gesture Recognition dataset



Scene recognition dataset



Daily activities dataset

[<http://link.springer.com/article/10.1007/s11042-016-3374-6>]
[<http://www0.cs.ucl.ac.uk/staff/M.Firman/RGBDdatasets>]

III. Dispozitive de recunoaștere a gesturilor

Consola Kinect - Resurse

Download

- <https://www.microsoft.com/en-us/download/details.aspx?id=44561>

Kinect Programming Walkthroughs

- <http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/guides.aspx>

Coding4Fun Kinect Toolkit

- <http://c4fkinect.codeplex.com>

Kinect SDK Quickstarts

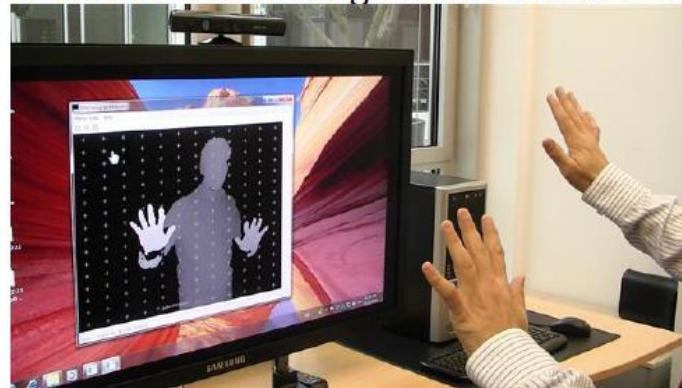
- <http://channel9.msdn.com/series/KinectSDKQuickstarts>

III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

IPLAB & Kinect

- Usage of the Kinect System for homework and student's projects
 - (Computer Vision course - Prof. S. Battiato - University of Catania)
- 2 Research projects:
 1. POR4.1.1.2 EMOCUBE (EdisonWeb)
 2. POR4.1.1.1 DIGINTEGRA (EdisonWeb, CCR,...)
 - Analyze the feedback of audiovisual advertising
 - Integrate of interactive multimedia content through natural interface



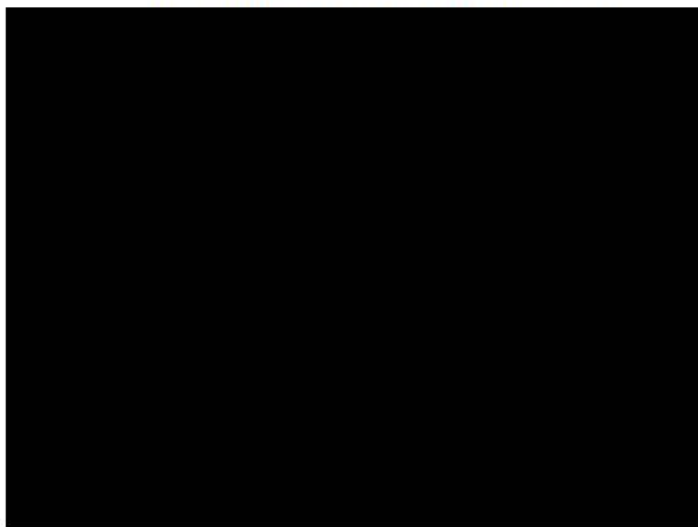
[<http://unict.it/>]

III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

Virtual Dressing Room

- The virtual fitting room is a program which allows users to select and try on clothing all from their Kinect.



<http://www.youtube.com/watch?v=1jbvnk1T4vQ>

[\[http://unict.it/\]](http://unict.it/)

III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect



III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

KinectFusion: Merges Real and Virtual Worlds

- Generate three-dimensional (3D) models in real time using a standard Kinect system.
- The technology can scan in 3D:
 - Objects
 - People
 - Entire rooms



III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

DepthJS

- DepthJS is a web browser extension that allows any web page to interact with the Microsoft Kinect via Javascript.



III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

Kinect in Hospitals

- Kinect also shows compelling potential for use in medicine
- Kinect for intraoperative, review of medical imaging, allowing the surgeon to access the information without contamination



III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

Etronika – online banking



III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

Etronika – online banking

Autentificare

- voce
- față
- cod PIN

Informații afișate

- informații despre carduri / credite / conturi / extrase / cursuri

Operații

- schimb valutar
- mutare între conturi proprii
- plăți după anumite template-uri

[<http://channel9.msdn.com/coding4fun/kinect/Banking-Kinect-Style>]

III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

Alte idei de aplicații

<http://www.openni.org/showcase/>

III. Dispozitive de recunoaștere a gesturilor

Aplicații Kinect

- Se pot implementa o multitudine de idei utilizând facilitățile dispozitivului kinect
- **Ce avem nevoie**
 - Cunoștințe de C++ / C#
 - Cunoștințe de Machine Learning și Computer Vision
 - WPF (util și dacă dorim să dezvoltăm o aplicație Windows Phone) + XNA (pentru controlul dispozitivelor periferice)

Concluzii

- Tehnologiile de detecție automată a gesturilor sunt într-o continua dezvoltare;
- Dispozitive de recunoaștere a gesturilor
 - Dispozitive cu contact;
 - Dispozitive fără contact:
 - Camere de adâncime;
 - Consola Kinect;
 - Aplicații Kinect.



Întrebări?

Ce urmează?

Componentele unui sistem de recunoaștere

