

8. Interfețe bazate pe senzori de adâncime

Senzori de adâncime existenți

Dintre tehnicile de măsurare a adâncimii obiectelor / scenei (fig. 1) abordările bazate pe senzori optici sunt singurele care oferă posibilitatea obținerii unor hărți de adâncime cu rezoluție ridicată (dense).

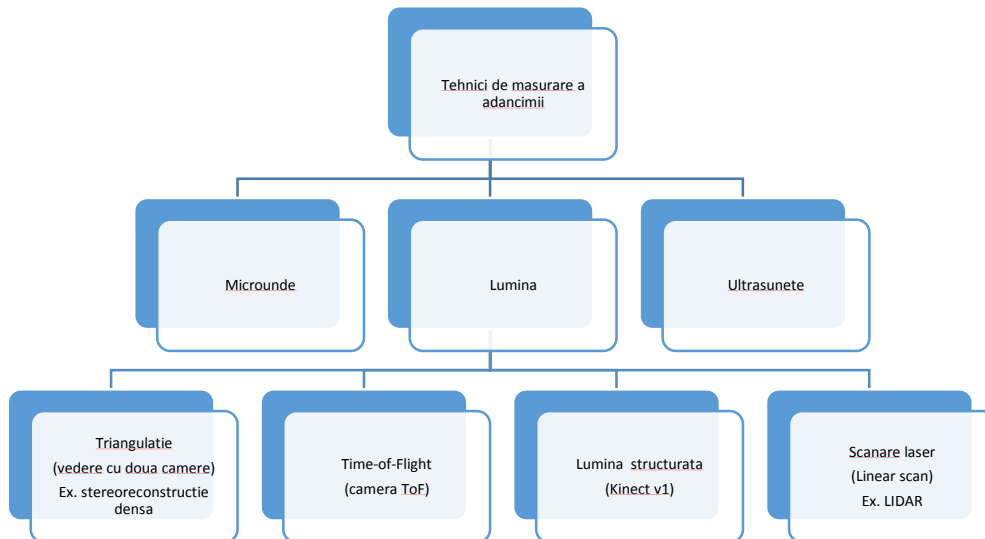


Fig. 8.1. Ierarhizarea tehnicilor de măsurare a adâncimii.

Senzorii de stereo-viziune permit obținerea unei hărți de adâncime dense prin corelația proiecțiilor (razelor de lumină reflectate) ale punctelor din scenă pe cele două plane imagine [Nedevschi2012]. *Avantajul* unui astfel de sensor este **natura pasivă** (captează lumina ambientală reflectată de pe obiectele din scenă) și rezoluția hărții de adâncime generate (egală cu rezoluția imaginii 2D) concomitent cu distanța mare de detecție (depinde în principal de distanța dintre camere: *baseline*) și posibilitatea de utilizare în scenarii outdoor (<https://cv.utcluj.ro/index.php/intersafe-2.html>). Ca principal *dezavantaj* amintim efortul de calcul necesar etapelor implicate în reconstrucție: rectificarea imaginilor și stereo-corelația. Implementarea acestor algoritmi prin soluții hardware [Pantilie2011] reduce timpul de procesare dar crește costurile sistemului în condițiile în care costurile unor camere profesionale nu sunt neglijabile. Cum majoritatea camerelor de luat vederi funcționează doar în spectrul vizibil de lumină, utilizarea sistemelor stereo în condiții de vizibilitate redusă este limitată.

Celelalte categorii de **senzori optici sunt de tip activ** deoarece se bazează pe captarea imaginii reflectate de pe obiectele din scenă a unei surse de lumină (de obicei în spectrul IR apropiat).

Scanerile laser se bazează pe măsurarea timpului de propagare al unui spot laser între momentul emisiei de către sursă și momentul recepției de către receptor. Scanarea scenei este liniară, lumina emisă de fiecare emițător fiind proiectată circular prin intermediul unei oglinzi cu un câmp vizual angular (azimut) de 360° și o rezoluție angulară dependentă de producător. Prin combinarea mai multor emițătoare se poate

mări câmpul vizual vertical (elevația) și implicit rezoluția verticală a senzorului, putându-se ajunge la obținerea unei hărți de adâncime cvasi-dense ca și în cazul Velodyne-LIDAR [<https://velodynelidar.com/products/>], [Schwarz2010]. Ca și principal *avantaj* amintim acuratețea foarte bună a măsurătorilor 3D, câmpul vizual orizontal (azimut) de 360°, distanța mare de detecție (putând fi utilizate în scenarii outdoor). Ca *dezavantaj* amintim faptul că rezoluția spațială este scăzută (depinde de numărul emițătoarelor laser), mărirea numărului de emițătoare fiind proporțional cu costul echipamentului care oricum este foarte ridicat (Velodyne lidar: 8.000 ... 70.000 USD).

VELODYNE

<http://velodynelidar.com/lidar/lidar.aspx>

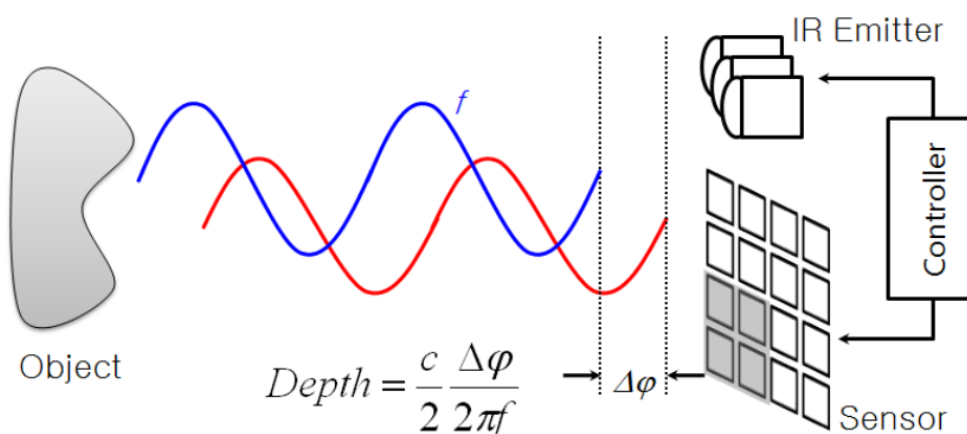
3D model of New Orleans based on Nokia Here LIDAR data:

<https://www.youtube.com/watch?v=75yJUW91ITs>

Zoox: <https://www.youtube.com/watch?v=fZfeSY2hrY4>,

<https://www.youtube.com/watch?v=xgXH4y0s5Ho>

Camerele de adâncime de tip **Time-of-Flight (ToF)** se bazează pe iluminarea scenei cu lumină IR și măsurarea distanței pe baza timpului de propagare a razelor de lumină reflectate de pe obiectele din scenă. Din punctul de vedere al principiului de funcționare există 2 tipuri de camere ToF: *cu impulsuri luminoase și cu modulația în amplitudine a undelor de lumină (continue)*. Primele se confruntă cu problematica măsurării unor intervale de timp foarte scurte pentru a obține rezoluții de adâncime de ordinul centimetrilor. Cele bazate pe modulația undelor luminoase continue măsoară deplasarea de fază dintre raza de lumină modulată emisă și cea recepționată, ceea ce corespunde cu timpul de propagare (ToF) și implicit cu adâncimea punctului respectiv. Totuși și în acest caz pot apărea ambiguități de ordinul multiplilor lungimii de undă modulate. Principalele *dezavantaje* ale camerelor de tip ToF sunt rezoluția scăzută pentru camerele existente pe piață și sensibilitatea scăzută (implicit nivel ridicat de zgomot). Ca și *avantaj* față de camerele de adâncime bazate pe lumina structurată este acuratețea mai bună a reconstrucției 3D [Langmann2012]. Ca domeniul de utilizare se pot folosi doar în scenarii de tip indoor și distanțe limitate (de ordinul metrilor). Există și dezvoltări de metode care să suprimă influența luminii ambientale (ex. *Suppression of Background Illumination - SBI*).



Principiul de masurare al distantei la camerele de tip TOF [Hansard2012]

Camerele de adâncime bazate pe **lumină structurată** au în componență un emițător laser IR și o cameră sensibilă la spectrul IR (senzor de imagine monocrom acoperit cu un filtru IR). Emițătorul laser emite o rază de lumină în IR care este divizată apoi prin intermediul unei rețele de difracție pentru a crea un șablon de puncte (pete/pistrui / „*spekles*”) pseudo-aleator care este proiectat pe scenă (fig. 8.2). Pe baza acestui model, pentru fiecare vecinătate a unui punct din imaginea proiectată se obține un „cuvânt-cod” care etichetează acel punct în mod unic. Imaginea reflectată a scenei este captată de camera IR și corelată cu un șablon de referință. Șablonul de referință este obținut prin captarea imaginii reflectate de pe un plan plasat la o distanță cunoscută față de senzor. Când un punct (*speckle*) este proiectat pe un obiect aflat la o distanță mai mare sau mai mică decât planul de referință, poziția punctului din imaginea IR reflectată este deplasată pe direcția de variație a disparității (direcție paralelă cu *baseline*-ul sistemului) până când se obține cea mai bună potrivire dintre imaginea reflectată și imaginea de referință. Ca și metrică de potrivire se folosește o funcție de corelație simplă calculată pe vecinătățile punctelor candidat. Prin calculul acestor deplasamente pentru fiecare punct din imaginea IR a scenei se obține o hartă de disparități din care se estimează adâncimea fiecărui punct [[Khoshelham2012](#)], cunoscând parametrii intrinseci ai camerei IR și poziția și orientarea relativă dintre camera IR și proiectorul IR.



Fig. 8.2. Șablonul de puncte (pete/pistrui/„*spekles*”) proiectat pe scenă în cazul unui senzor de adâncime cu lumină structurată.

Un exemplu de cameră de adâncime bazat pe lumina structurată este senzorul Kinect (versiunea 1). Ca și principal avantaj al acestui senzor este prețul scăzut și disponibilitatea ca și bun de larg consum împreună cu librării de dezvoltare puse la dispoziție de producător (Microsoft). Ca și dezavantaje menționăm limitele reduse ale detecție (maxim 4.5m) și sensibilitatea la lumina ambientală, utilizarea lui pretându-se la scenarii indoor (Kinect v1).

Tabel 8.1. Sinteză comparativă a senzorilor de adâncime prezentați

Caracteristici	Stereo viziune	Lumină Structurată	Time of Flight (ToF)	Scanare laser (LIDAR)
Efort de calcul	Mare - implem. soft Mic - implem. hard	-	-	-
Cost echipament	Mediu - soluții profesionale (~10.000 \$)	Scăzut (Kinect v1 – 200 \$)	Mediu - camere profesionale (~10.000 \$) Scăzut (Kinect v2 - 200\$)	Ridicat ~ rezoluția spațială (ex. Velodyne: 8.000 .. 70.000 \$)
Rezoluție spațială	Personalizabilă în funcție de aplicație	Mică (Kinect v1)	Mică – soluții profesionale Medie (Kinect v2)	Mică
Acuratețea reconstrucție	~ acuratețea parametrilor de calibrare și metoda de stereo-corelație	Medie	Bună	Foarte bună
Domeniu de distanțe	Orice domeniu de distanțe (.. ~100 m) ~ parametrii sistemului	Apropiat (~ m)	Apropiat (~ m)	Orice domeniu de distanțe (.. ~100 m)
Utilizare	Indoor/ Outdoor Limitare în condiții de vizibilitate scăzute (întuneric, ceață, ninsoare, ploaie)	Indoor	Indoor / Outdoor (SBI)	Indoor/ Outdoor Limitare în condiții atmosferice nefavorabile

Senzorul Kinect

Caracteristici hardware

Microsoft Kinect [Kean2011] este un dispozitiv dezvoltat inițial pentru urmărirea (*tracking*-ul) mișcărilor corpului uman cu aplicații în jocuri. Dispozitivul hardware este însoțit de un kit de dezvoltare software care prin facilitățile oferite devine foarte util în aplicațiile care implică metode interactive de tip “*hands-free*” în care interacțiunea cu calculatorul se poate face prin intermediul gesturilor. Dar versatilitatea

sistemului face ca utilizarea lui să poată fi extinsă și la alte domenii: sisteme de securitate / biometrice, medicină sau aplicații industriale.

Prima versiune a dispozitivului Kinect (v1), lansată în anul 2010 are la bază tehnologia dezvoltată de firma PrimeSense (fig. 8.3.a). Dispozitivul este compus dintr-o cameră RGB, un senzor de adâncime (emițător IR și cameră IR) și o arie de microfoane pentru recunoașterea vocii și a sunetelor în general, un motor de înclinare („*motorized tilt*”) pentru înclinarea camerei cu 30 grade în sus sau în jos și un accelerometru pe 3 axe (fig. 8.3.b).

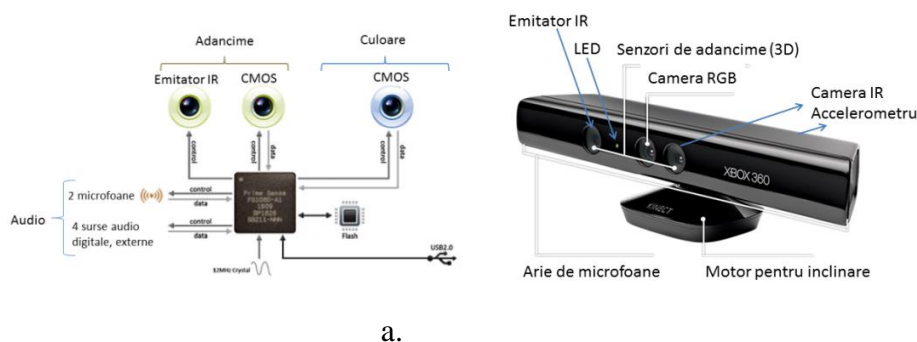


Fig. 8.3. Structura hardware (a) și vedere de ansamblu (b) a senzorului Kinect (v1).

Pentru măsurarea adâncimii Kinect-v1 folosește principiul luminii structurate. Un emițător (dioda laser IR) generează șablonul de puncte pseudo-aleatoriu (*speckles*) prin intermediul unei rețele de difracție [Kinect2011]. Imaginea reflectată a șablonului de lumină structurată este captată de o cameră CMOS în IR cu rezoluția spațială de 320x240 pixeli și distanța focală de 6 mm. Prin corelarea șablonului generat de emițătorul IR cu imaginea lui reflectată se generează o hartă de adâncime densă (figura 8.4) cu o rezoluție de adâncime codificată pe 11 biți (2^{11} valori discrete de adâncime ~ 7 cm la distanța maximă de detecție) și cu o eroare de estimare a adâncimii de cca 1% la o distanță maximă de detecție (5 m) [Khoshelham2012].

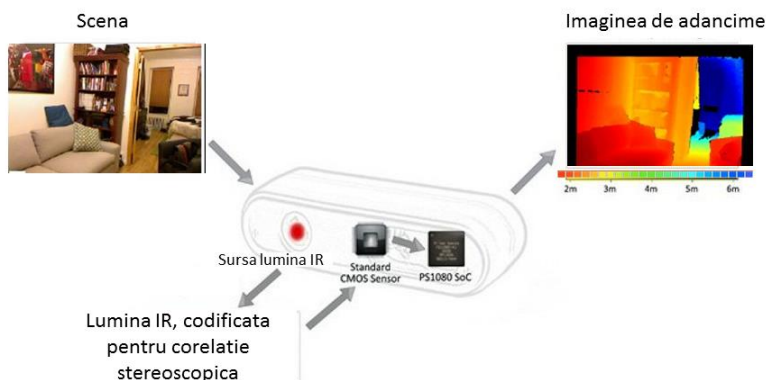


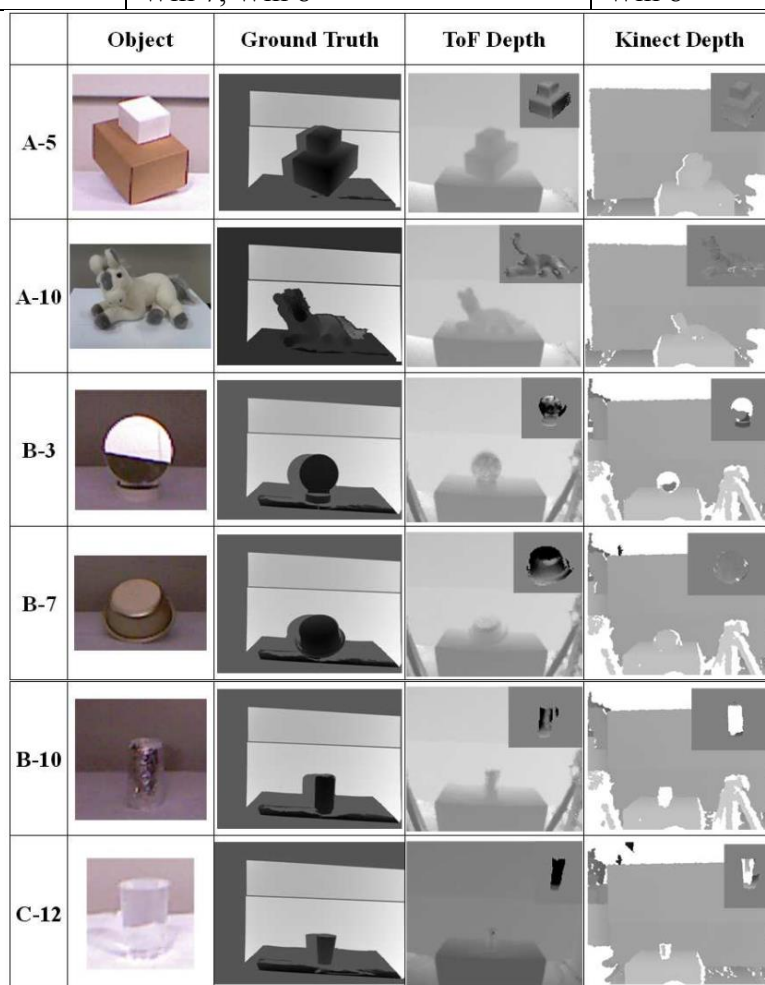
Fig. 8.4. Imaginea de adâncime cu senzorului Kinect (v1).

Versiunea v2 a dispozitivului Kinect [Kinect2014] lansată în 2014 aduce o serie de îmbunătățiri dintre care principală este schimbarea tehnologiei de măsurare a adâncimii: *Time-of-Flight* (ToF) și folosirea unei rezoluții spațiale pentru imaginea de adâncime mai mari decât la camerele de tip ToF existente pe piață

[Langmann2012]. O descriere comparativă a caracteristicilor celor două versiuni ale senzorului Kinect este prezentată în tabelul 8.2:

Tabel 8.2. Comparație Kinect-v1 vs. Kinect-v2

Trăsătură	Kinect v1 (2010)	Kinect v2 (2014)
Chip	PrimeSense	N/A
Senzor color	640 x 480 @30 fps	1920 x 1080 @30 fps
Câmp vizual (°)	62 x 48.6	84.1 x 53.8
Senzor de adâncime	Lumină structurată	Time-of-Flight (ToF)
Rezoluție hartă adâncime	320 x 240	512 x 424
Adâncime maximă	~4.5 M	~4.5 M
Adâncime minimă	40 cm (în modul „near”)	50 cm
Câmp vizual (°)	58.5 x 46.6	70.6 x 60
Alte trăsături		
Reglaj motorizat înclinare	DA	NU
Nr. articulații scheleton	20	25
Nr. scheletoane urmăribile	2	6
Modelarea fețelor	116 noduri	1000 noduri
Detectie mana/degete	-	5 stari: opened/closed/lasoo/not tracked, unknown
Interfață USB	2.0	3.0
Sisteme de operare	Win 7, Win 8	Win 8



Evaluare calitativa comparativa a reconstrucției 3D TOF vs. lumina structurata (Kinect). Calitatea depinde de reflectivitatea suprafetelor obiectelor: A – materiale cu suprafata difuza; B – materiale cu suprafata speculara; C – materiale cu supafata translucida [Hansard2012].

Aplicații ale senzorului Kinect

Cea mai frecventă utilizare a senzorului Kinect este în dezvoltarea de aplicații interpretare a gesturilor. Acest gen de aplicații se bazează pe urmărirea poziției încheieturilor scheletonului și inferarea gesturilor prin diferiți algoritmi (DTW – *Dynamic Time Warping* [Senin2013][Reyes2011] sau metode de clasificare [Patsadu2012]). De asemenea se pot dezvolta algoritmi de detecție și urmărire a degetelor și inferare a gesturilor făcute cu acestea [Cerezo2012] [Oikonomidis2011]. Din aceasta categorie de aplicații amintim:

- Garderoba virtuală – permite utilizatorului să probeze virtual haine [KinectApp1]
- Interacțiunea cu calculatorul prin gesturi – *DepthJS* este o aplicație dezvoltată de MIT Media Lab ce permite interacțiunea cu orice browser web prin intermediul gesturilor [KinectApp2]. De asemenea se pot dezvolta aplicații de interpretare ale gesturilor în domeniul medical (de ex. în chirurgie pentru vizualizarea/navigarea în colecțiile de imagini medicale în timpul operației) .
- Controlul unor aparate prin gesturi – permite controlul unui aparat de zbor de tip *quadrotor* prin gesturile mâinii [KinectApp3]
- Implementarea unor instrumente muzicale virtuale – permite cântatul la un instrument virtual (ex. pian) prin detecția poziției degetelor și a mișcării acestora [KinectApp4]
- Procesul de învățare a anumitor procedee medicale - aplicația *Mini-VREM (Mini-Virtual Reality Enhanced Mannequin)* poate fi folosit de Kinect pentru a crea un sistem care urmărește poziția mâinii, plasarea, rata și puterea de compresie pentru a învăța utilizatorii cum să facă resuscitare cardio-pulmonară în mod corect [KinectApp5]

De asemenea dispozitivul Kinect poate fi folosit ca și senzor de percepție mediului pentru vehicule autonome (în domeniul apropiat de distanțe de detecție). Un astfel de exemplu este folosirea lui pe un aparat de zbor de tip *quadrotor* pentru detecția și evitarea obstacolelor [KinectApp6].

Nu în ultimul rând poate fi folosit în sisteme de asistență a conducerii autovehiculelor pentru estimarea gradului de atenție sau oboseală a conducătorului auto [Werber2014] sau a gradului de conștientizare al șoferului înainte de a se urca la volan [Turpen2013].

Bibliografie

[Nedevschi2012] S. Nedevschi, R. Danescu, F. Oniga, T. Marita, *Tehnici de viziune artificială aplicate în conducerea automată a autovehiculelor*, Editura U.T. Press, Cluj-Napoca, 2012

[Hansard2012] Miles Hansard, Seungkyu Lee, Ouk Choi, Radu Horaud. *Time of Flight Cameras: Principles, Methods, and Applications*. Springer, pp.95, 2012, SpringerBriefs in Computer Science, ISBN 978-1-4471-4658-2

[Pantilie2011] C. D. Pantilie, I. Haller, M. Drulea, and S. Nedevschi, “[Real-Time Image Rectification and Stereo Reconstruction System on the GPU](#)”, in 2011 10th International Symposium on Parallel and Distributed Computing, 2011, p. 79–85.

[Velodyne2007] Velodyne Acoustics Inc., Velodyne’s Hdl-64e: A High Definition Lidar Sensor For 3-D Applications, White Paper, 2007, https://velodynelidar.com/downloads/#white_papers_first (citedt july 2020).

- [Schwarz2010] B. Schwarz, *Mapping the world in 3D*, Nature Photonics, Vol.4, p. 429-430, (2010).
- [Langmann2012] Langmann, B., Hartmann, K., & Loffeld, O. (2012, February). Depth Camera Technology Comparison and Performance Evaluation. In ICPRAM (2) (pp. 438-444).
- [Khoshelham2012] Khoshelham, K., Elberink, S. O. (2012). Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors (Basel, Switzerland)*, 12(2), 1437–1454.
<http://www.mdpi.com/1424-8220/12/2/1437>
- [Kinect2011] Kinect Pattern Uncovered, <https://azttm.wordpress.com/2011/04/03/kinect-pattern-uncovered/>
- [Senin2013] P. Senin, *Dynamic time warping algorithm review*, Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, p. 1-23. (cited Jan. 2013).
- [Patsadu2012] O. Patsadu, *Human gesture recognition using Kinect camera*, 2012 International Joint Conference on Computer Science and Software Engineering (JCSSE), p. 28–32 (2012).
- [Oikonomidis2011] I. Oikonomidis, N. Kyriazis, A.A. Argyros, *Efficient model-based 3D tracking of hand articulations using Kinect*, in Proceedings of the 22nd British Machine Vision Conference, BMVC'2011, (2011).
- [Cerezo2012] F. Trapero Cerezo, *3D Hand and Finger Recognition using Kinect*, Technical Report, Universidad de Granada (UGR), Spain, (2012).
- [Reyes2011] Reyes, M., Dominguez, G., Escalera, S. (2011, November). Feature weighting in dynamic timewarping for gesture recognition in depth data. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 1182-1188). IEEE.
- [Kinect2014] Microsoft Kinect for Windows, <http://www.microsoft.com/en-us/kinectforwindows/discover/features.aspx>, (citat iunie 2014)
- [KinectApp1] *Virtual Dressing Room*, <https://www.youtube.com/watch?v=1jbnk1T4vQ> (citat febr. 2013).
- [KinectApp2] *DepthJS*, MIT Media Lab, <http://depthjs.media.mit.edu/> (citat febr. 2013).
- [KinectApp3] *Interaction With A Quadrotor Via The Kinect*, ETH Zurich, <https://www.youtube.com/watch?v=A52FqfOi0Ek> (citat febr. 2013).
- [KinectApp4] *Kinect Fun Labs Election: Keyboard Anywhere*, <https://www.youtube.com/watch?v=4STUGl-YHDc>, (citat febr. 2013).
- [KinectApp5] *Mini-VREM (Mini-Virtual Reality Enhanced Mannequin)*, <http://www.mini-vrem.it/> (citat febr. 2013).
- [KinectApp6] *Quadrotor Autonomous Flight and Obstacle Avoidance with Kinect Sensor*, <http://hybrid.eecs.berkeley.edu/~bouffard/kinect.html>, (citat febr. 2013).
- [Turpen2013] A. Turpen, *Toyota using Driver Awareness Research Vehicle with Microsoft Kinect to combat distracted driving*, nov. 2013, http://www.carnewscafe.com/2013/11/toyota_darv_microsoft_kinect/, (citat Apr. 2015)
- [Werber2014] L. Li, K. Werber, C. F. Calvillo, K. D. Dinh, A. Guardé, A. König, *Multi-sensor soft-computing system for driver drowsiness detection*, In *Soft Computing in Industrial Applications*, Springer International Publishing, p. 129-140, (2014).
http://dap.vsb.cz/wsc17conf/Media/Default/Page/online_wsc17_submission_40.pdf

Senzorii Intel RealSense

<https://software.intel.com/en-us/realsense/devkit>

Intel® RealSense™ 3D Camera (F200) este destinat pentru urmarirea capului si mainilor in 3D (fiind indreptat catre user), in timp ce Intel® RealSense™ 3D Camera (R200) a fost dezvoltat pentru a urmari mediul in 3D (fiind indreptat in partea opusa user-ului).

Tabel 8.3. Specificatii hardware F200 - R200

Feature	Intel® RealSense™ 3D Camera (F200)	Intel® RealSense™ 3D Camera (R200)
Target Platform	All-in-Ones, Notebooks, 2-in-1s	Tablets, 2-in-1s, and more
Range	Indoor (0.2 meters - 1.2 meters, indoors only) / structured light	Indoor: 0.5 .. 3.5 m Outdoor: 10m (IR stereo)
Depth/IR	640x480 resolution at 60fps	640x480 resolution at 60fps
RGB	1080p at 30fps	1080p at 30fps
USB	USB 3.0 required	USB 3.0 required
Supported OS	Supports Microsoft Windows* 8.1 (64-bit)	Supports Microsoft Windows* 8.1 (64-bit)
Developer Kit Dimensions	150mm x 30mm x 58mm	130mm x 20mm x 7mm
Price	99\$	99\$

F200: <https://software.intel.com/en-us/blogs/2015/01/26/can-your-webcam-do-this>

R200: <https://software.intel.com/en-us/articles/realsense-r200-camera>

Tabel 8.4. Capabilitati software F200 - R200

Intel® RealSense™ 3D Camera (F200)	Intel® RealSense™ 3D Camera (R200)
Full hand-skeletal tracking and gesture control	Measurement
3D segmentation	Depth-enhanced photography and video
Facial analysis	Face detection and tracking
Depth-enhanced Augmented Reality	Depth-enhanced Augmented Reality
Speech	
3D Capture for faces	3D Capture: faces, people, and environments

<https://software.intel.com/en-us/articles/intel-realsense-data-ranges> :

<http://roscon.ros.org/2015/presentations/ROSCon-2015-Intel-RealSense-Robotics-Innovation-Program-ROS-interface.pdf>

Intel RealSense SDK: Hand Tracking module [F200]

The SDK hand tracking module provides **real-time 3D hand motion tracking**, using a single depth sensor. The hand module can track **one or two hands**, providing precise joint-level locations and positions. The module can also identify “**gestures**”, which are certain significant hand postures or motions, for example **a wave, tap or thumbs-up sign**.

Using the hand module, you can enable your application to be controlled by hand motions, using visual cues alone (without a touch interface). For instance, you can interpret a hand tap as selection, a hand swipe as scrolling, and so on.

Tracking Modes: The hand module has two main tracking modes, which differ by the information they provide and the computation resources that they require:

- **Full-hand** - returns the **full 3D skeleton of the hand, including all 22 joints, fingers information, gestures**, and more.
- **Extremities** – returns the general location of the hand, its silhouette, and the extremities of the hand (6 points): the hand’s top-most, bottom-most, right-most, left-most, center and closest (to the sensor) points).

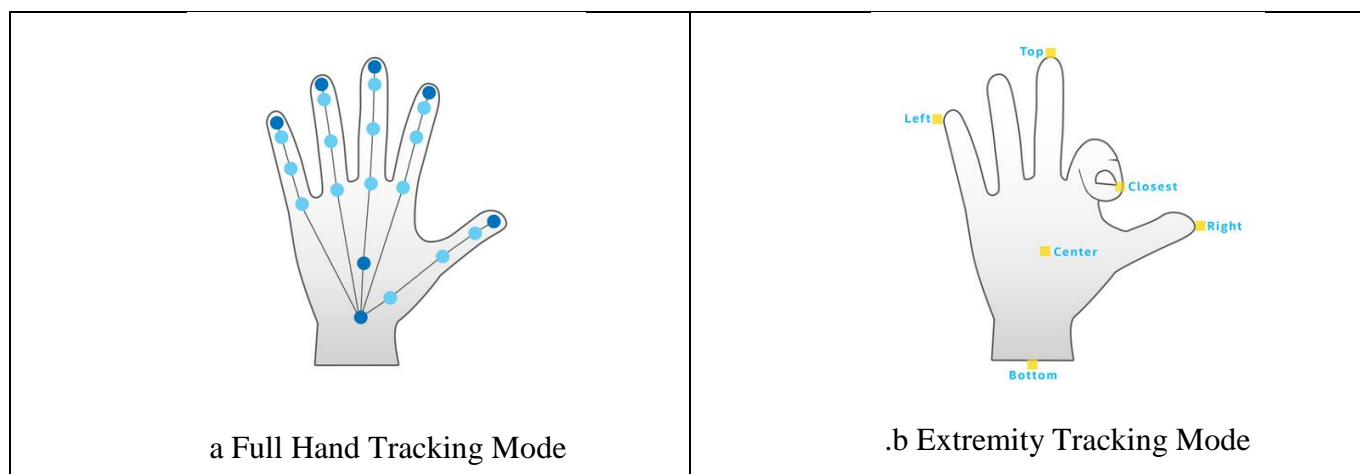


Fig. 8.5. a Full Hand Tracking Mode; b. Extremity Tracking Mode

Use the full-hand tracking mode when you need to access to specific skeletal information. Use the extremities tracking mode when you only need to access the entire hand location. The extremities mode conserves resources such as CPU, power, computation threads and memory. Depending on the requirements of your application, you may wish to use the blob module instead. Table 12 below summarizes the three tracking options:

Tracking Mode	Hand only?	Outputs	Computational resources	Limitations
Full Hand	Yes	Segmentation image, extremity points, hand side, alerts, joints info, fingers info, openness, gestures	Highest, multiple threads	2 hands, 60 cm range, slower hand speed
Extremities	Yes	Segmentation image, extremity points, hand side, alerts	Medium, single thread	2 hands, 60 cm range, medium hand speed
Blob	No	Segmentation image, extremity points, contour line	Low, single thread	4 objects, 100 cm range, fast speed

Table 12: Tracking Options

The following types of data are available only when in full-hand tracking mode:

- Joint information
- Finger data
- Hand side
- Openness
- Gestures

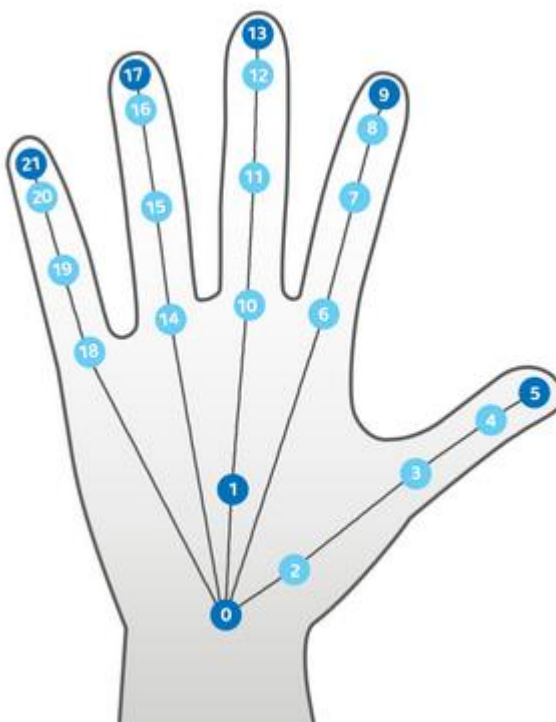
Hand Joints

The wrist joint and all the finger joints correspond to actual joints in the hand (where two bones connect). The palm-center joint is a “virtual joint” that corresponds to the center of the palm.

Note that the joints are arranged in a tree-like structure, where each joint has a “parent” joint, and the root is the wrist joint. The parent-child relationships are used to express joint rotation.

#	Joint Name	Joint Parent	#	Joint Name	Joint Parent
0	JOINT_WRIST	NA	11	JOINT_MIDDLE_JT1	10
1	JOINT_CENTER	0	12	JOINT_MIDDLE_JT2	11
2	JOINT_THUMB_BASE	0	13	JOINT_MIDDLE_TIP	12
3	JOINT_THUMB_JT1	2	14	JOINT_RING_BASE	0
4	JOINT_THUMB_JT2	3	15	JOINT_RING_JT1	14
5	JOINT_THUMB_TIP	4	16	JOINT_RING_JT2	15
6	JOINT_INDEX_BASE	0	17	JOINT_RING_TIP	16
7	JOINT_INDEX_JT1	6	18	JOINT_PINKY_BASE	0
8	JOINT_INDEX_JT2	7	19	JOINT_PINKY_JT1	18
9	JOINT_INDEX_TIP	8	20	JOINT_PINKY_JT2	19
10	JOINT_MIDDLE_BASE	1	21	JOINT_PINKY_TIP	20

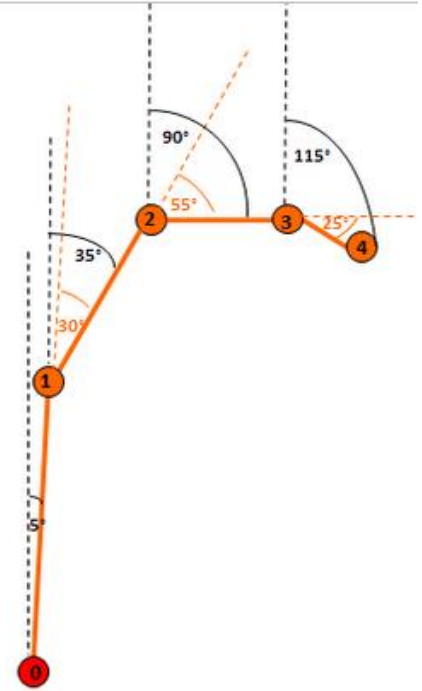
Table 13: The Joint Name and Index.



Local and Global Joint Rotation

A joint's rotation refers to the angle of the joint's bone, which is the bone that ends at the joint and starts at the joint's parent. The hand module expresses joint rotation in two ways:

- **Local rotation** - the angle between the joint's bone and its parent's bone. This value is stored in the localRotation property of [JointData](#)
- **Global rotation** - the angle between the joint's bone and the camera coordinate system (Y axis). This value is stored in the globalOrientation property of [JointData](#)









Handling Gestures







Gestures are specific hand motions or postures. It is often convenient to have your application respond only to pre-defined gestures, rather than having to process raw position data.



Table 15 summarizes the gestures supported in the default gesture pack. Certain pairs of gestures are similar, and gesture detection can confuse them. Therefore, it's advisable not to enable both gestures at once. Table 15 also summarizes the restrictions. In general, it is best practice to enable only those gestures that are relevant to the specific stage in your application, rather than enabling all gestures at all times. At the beginning of each stage, enable the required gestures and disable the gestures that are not required.

The Gesture Table

Gesture	Illustration	Don't Enable with:	Description

click		two_fingers_pinc h_open	Open hand facing the camera, moves the index finger quickly toward the palm center.
fist		full_pinch	All fingers folded into a fist. The fist can be in different orientations as long as the palm is in the general direction of the camera.
full_pinch		fist	All fingers extended and touching the thumb. The pinched fingers can be anywhere between pointing directly to the screen or in profile.
spreadfingers		NA	Hand open, facing the camera.
swipe_down		NA	Hand with palm facing the camera, moves down and immediately back to the starting position.
swipe_left		wave	Hand with palm facing the camera, moves left and immediately back to the starting position.

swipe_right		wave	Hand with palm facing the camera, moves right and immediately back to the starting position.
swipe_up		NA	Hand with palm facing the camera, moves up and immediately back to the starting position.
tap		NA	A hand in a natural relaxed pose is moved forward as if pressing a button.
thumb_down		NA	Hand closed with thumb pointing down.
thumb_up		NA	Hand closed with thumb pointing up.
two_fingers_pinch_open		click	Hand open with thumb and index finger touching each other.

v_sign		NA	Hand closed with index finger and middle finger pointing up.
wave		swipe_left swipe_right	An open hand facing the screen. The wave can include any number of repetitions.