

Ministerul Educaţiei și Cercetării al Republicii Moldova

Universitatea Tehnică a Moldovei

Departamentul Informatică și Ingineria Sistemelor

**Referat**

Sisteme de Vedere Artificială

Aplicații ale sistemelor de vedere artificială

A efectuat: Avxentiev Dorin

Gr.RM-211

A verificat: Cărbune Viorel

**Chişinău – 2024**

**Cuprins**

**Introducere.................................................................................... 3**

**Structura sistemele de vedere artificială .....................................4**

**Obiective ........................................................................................8**

**Clasificarea (Computer Vision) ...................................................9**

**Aplicații ale viziunii computerizate ...........................................10**

**Analiză comparativă între ochiul uman și o cameră ...............12**

**Modul de achiziție (Tipuri de senzori) ......................................14**

**Concluzia .....................................................................................16**

**Bibliografie ..................................................................................17**

**Introducere**

**Vederea artificială (computer vision)** este domeniul care se ocupă cu dezvoltarea de algoritmi și tehnici matematice pentru a permite calculatoarelor să perceapă și să interpreteze imagini sau secvențe video într-un mod similar vederii umane. Acest proces implică recuperarea informațiilor tridimensionale despre forma și aspectul obiectelor din imagini, recunoașterea obiectelor, a persoanelor și a scenelor, precum și estimarea poziției și mișcării lor. Deși tehnologiile moderne pot genera modele 3D precise și pot detecta anumite elemente din imagini, ele nu ating încă nivelul complexității și acurateței vederii umane.

Viziunea computerizată este domeniul inteligenței artificiale care s-a concentrat pe a permite computerelor și sistemelor să proceseze și să obțină informații din date vizuale, cum ar fi imagini și videoclipuri, așa cum ar face un om. Viziunea computerizată permite computerelor să analizeze intrarea vizuală la nivel de pixel și să o înțeleagă. Pentru înțelegerea mai buna a conceptului de vederea artificialăde obicei se face analogia acesteia cu viziunea umană. Ambele se bazează pe un principiu similar, însă, atunci când, oamenii, observăm lucruri, avem deja context pe baza căruia putem identifica obiectele.În vederea artificială, mașinile au nevoie de date și algoritmi pentru a învăța cum să îndeplinească aceste funcții. Diferența este că computerele pot analiza cantități mari de date și pot analiza nenumărate produse în fiecare unitate de timp, ceea ce oamenii, în mod natural, nu sunt capabili sa indeplineasca.



**Structura sistemele de vedere artificială**

Sistemele de vedere artificială (computer vision) sunt complexe și au o structură bine definită, care include mai multe componente interconectate. Aceste componente lucrează împreună pentru a capta, prelucra și interpreta informațiile vizuale. Principalele elemente din structura acestor sisteme sunt prezentate in fig. 1:

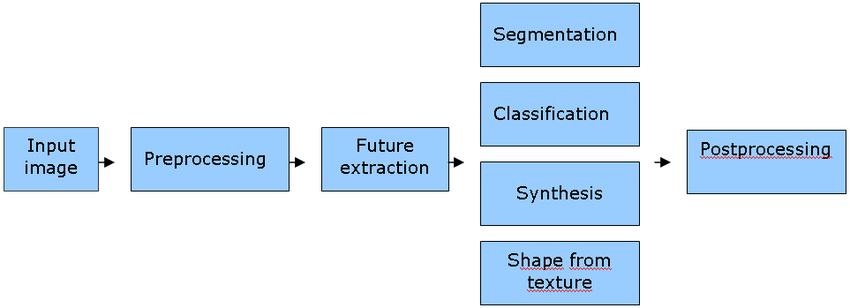


Fig.1 Structurasistemele de vedere artificială

* 1. **Achiziția imaginilor (Echipament de achizitie a imaginilor)**

**1.1.1 Senzori și camere**

În vederea artificială, imagini sau secvențe video sunt capturate folosind dispozitive de luat vederi, cum ar fi camere digitale, camere stereo sau camere de adâncime (depth cameras) precum Kinect. Camerele pot capta date vizuale fie în spectrul vizibil, fie în alte spectre (infraroșu, ultraviolete).

* + 1. **Preprocesarea datelor**

Imaginile brute sunt adesea preprocesate pentru a îmbunătăți calitatea (de exemplu, prin reducerea zgomotului sau ajustarea contrastului) și pentru a corecta distorsiunile optice cauzate de lentilele camerei. Metodele de preprocesare includ: transformari la nivelul de gri (Fig.2), transformari geometrice, preprocesari locale, refacerea imaginilor.

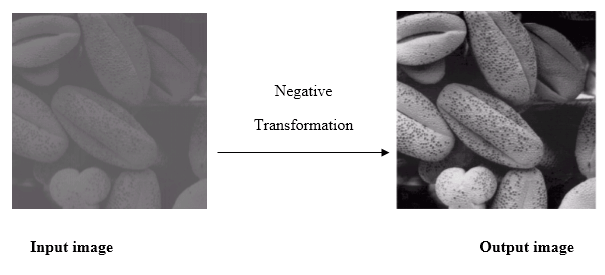


Fig.2 Exemplu transformării la nivel de gri.

* 1. **Procesarea și filtrarea imaginilor**

Imaginile sunt procesate pentru a corecta orientările sau pentru a ajusta dimensiunile, utilizând operații precum redimensionarea, rotația sau translația. Filtrarea implică aplicarea unor tehnici, cum ar fi filtrele Gaussian sau Sobel, pentru a extrage informații importante, precum marginile obiectelor sau texturile. Segmentarea imaginii constă în împărțirea acesteia în regiuni mai mici, bazate pe caracteristici comune precum culoarea, textura sau forma, având scopul de a izola obiectele de interes de fundal.



Fig.3 Exemplu de procesarea și filtrarea a imaginii.

**1.3 Extragerea caracteristicilor**

**1.3.1 Detectarea și descrierea caracteristicilor**

Aceasta implică identificarea punctelor de interes, marginilor, colțurilor sau altor trăsături distinctive din imagine. Algoritmi precum SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features) sau ORB (Oriented FAST and Rotated BRIEF) sunt folosiți pentru a identifica caracteristici unice.După ce caracteristicile sunt detectate, ele sunt transformate în descrieri matematice care pot fi folosite pentru analiză ulterioară.

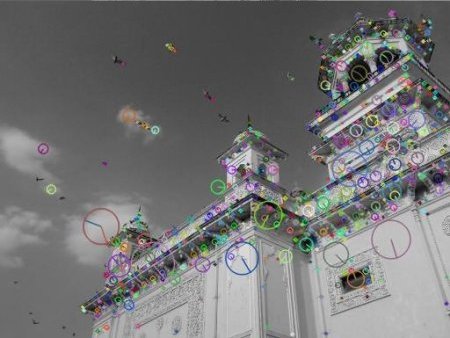


Fig.4 Scale-Invariant Feature Transform algoritm.

**1.3.2 Analiza și interpretarea obiectelor**

Utilizarea tehnicilor de clasificare și recunoaștere pentru a identifica obiectele din imagini. Acest lucru implică algoritmi de machine learning sau deep learning, cum ar fi rețele neuronale convoluționale (CNN) Fig.5, pentru a învăța dintr-un set de imagini etichetate și pentru a face predicții pe imagini noi. Reconstituirea 3D este bazat pe imagini bidimensionale, un sistem de vedere artificială poate încerca să reconstruiască forma tridimensională a unui obiect sau a unei scene, folosind tehnici precum stereo vision, structure from motion (SFM) sau time-of-flight (TOF).

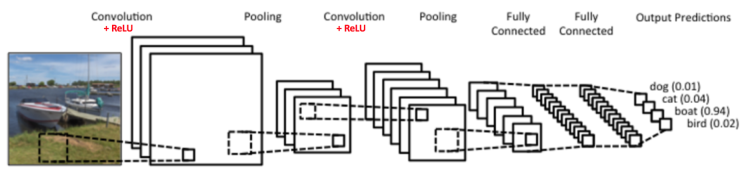


Fig.5 Rețele neuronale convoluționale.

**1.3.3 Urmărirea și detectarea mișcării**

Odată ce un obiect este detectat, sistemul poate urmări mișcarea acestuia în timp, utilizând algoritmi precum Kalman Filter Fig.6, sau algoritmi de tracking avansați.

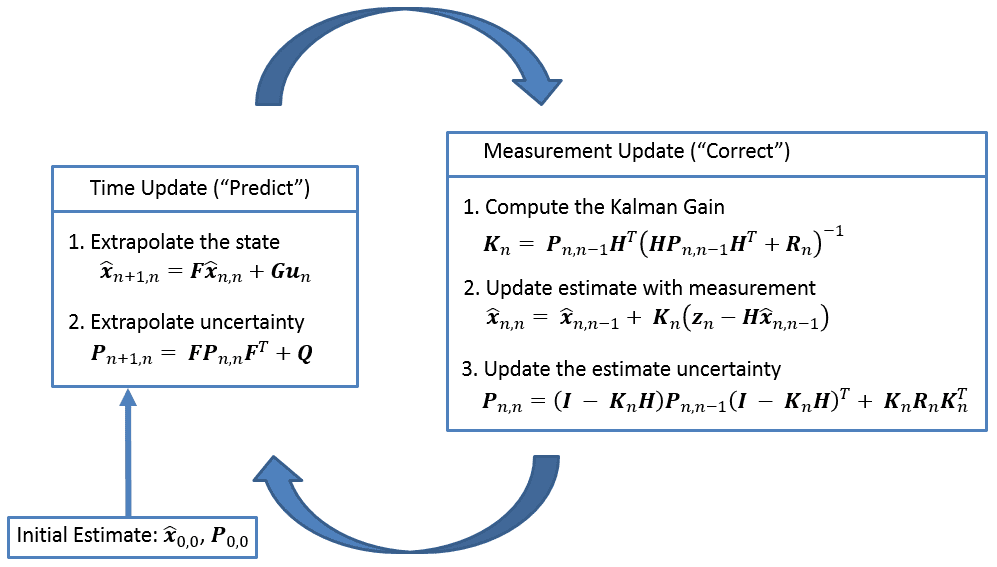


Fig.6 exemplu algoritmului Kalman Filter.

Filtrul Kalman este un algoritm care funcționează în două etape principale: **predicția** și **actualizarea**, repetând acest proces iterativ. Scopul său este să estimeze poziția și mișcarea unui obiect, folosind măsurători care pot fi zgomotoase sau incomplete.

**Predictia:** Se estimează starea viitoare a obiectului (poziție, viteză etc.) pe baza unui model matematic al mișcării (de exemplu, ecuațiile mișcării).

**Actualizarea:** Se primește o măsurătoare nouă (de exemplu, poziția detectată de un senzor).Diferența dintre predicție și măsurătoare (numită rezidual) este folosită pentru a corecta estimarea. Algoritmul combină predicția și măsurătoarea, acordând mai multă încredere sursei cu mai puțină incertitudine.

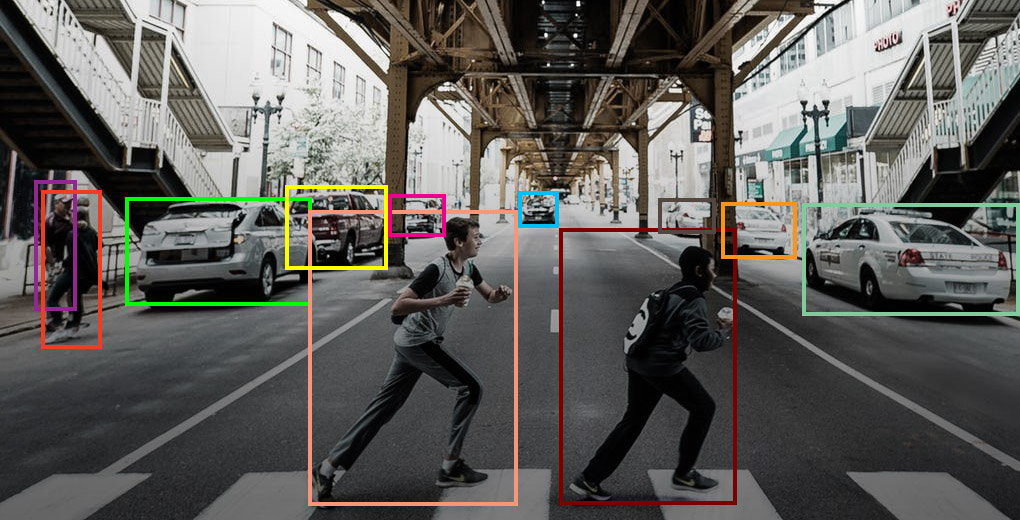


Fig .7 Detectarea miscării

**1.3.4 Luarea deciziilor și aplicații**

* **Aplicații în robotică:** Informațiile vizuale sunt folosite pentru a ghida mișcarea roboților, a-i ajuta să navigheze în medii complexe și să manipuleze obiecte.
* **Sisteme de recunoaștere facială:** Tehnicile de vedere artificială pot fi aplicate pentru a detecta și recunoaște fețele umane din imagini sau videoclipuri.
* **Analiza video:** Folosind vederea artificiala, sistemele pot analiza în timp real secvențele video pentru a detecta evenimente, activități sau anomalii.

**1.3.5 Stocarea și gestionarea datelor**

* **Baze de date și indexare:** Imaginile și informațiile extrase din ele sunt adesea stocate într-o bază de date pentru acces rapid și pentru a permite compararea ulterioară sau analiza.
* **Compresia și codarea:** Pentru eficiență, imaginile capturate sau procesate sunt deseori comprimate pentru a reduce dimensiunea de stocare și pentru a facilita transmiterea datelor.
* **Stocare distribuită:** Soluții precum Hadoop sau Amazon S3 permit stocarea masivă de date, necesare pentru proiecte la scară largă.

**Obiective**

Computer Vision urmărește dezvoltarea de tehnici și algoritmi care permit calculatoarelor să perceapă, interpreteze și înțeleagă lumea vizuală într-un mod asemănător oamenilor. Scopul principal este extragerea de informații semnificative din imagini și secvențe video, cum ar fi recunoașterea obiectelor, persoanelor, mișcărilor sau relațiilor spațiale, pentru a crea modele digitale ce imită percepția umană.

Metodele utilizate se bazează pe principii matematice și fizice, abordând o problemă inversă: deducerea structurii tridimensionale a lumii reale din date bidimensionale. Acest proces implică modele probabilistice și simularea percepției vizuale, inspirate de modul în care ochii sau camerele captează informația.Grafica joacă un rol complementar, permițând generarea imaginilor realiste prin simularea interacțiunii dintre lumină și suprafețe. Progresele în acest domeniu facilitează crearea de medii 3D detaliate și animarea personajelor, dar întâmpină provocări legate de replicarea completă a realismului uman.

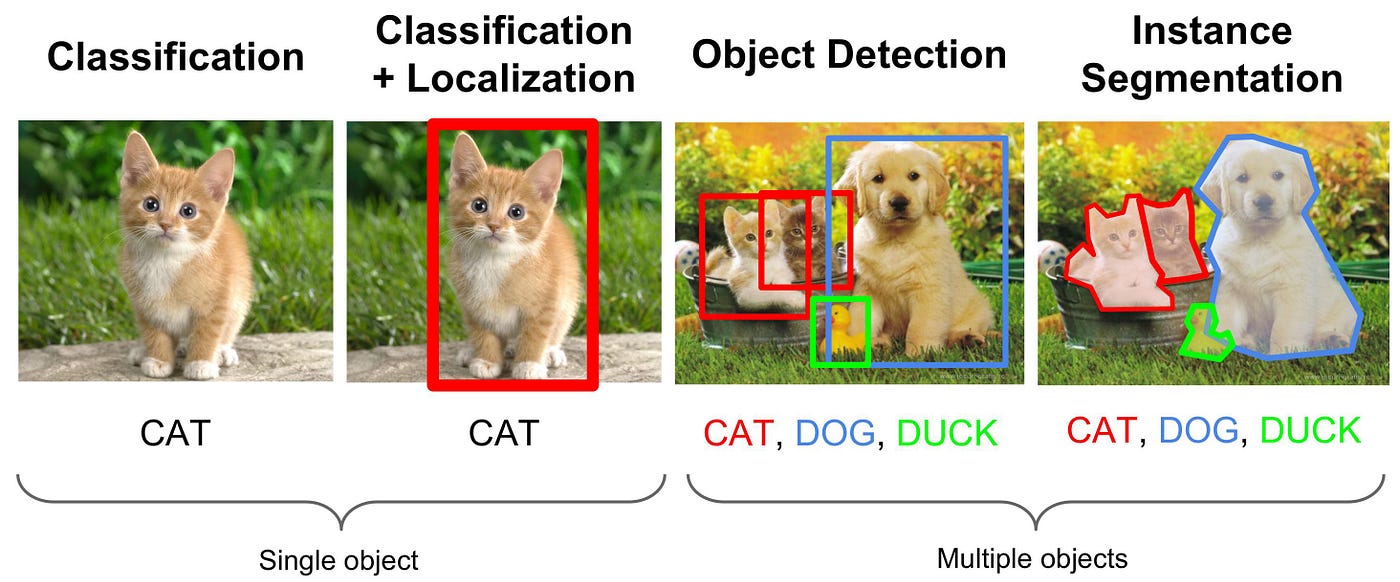


Fig.8 Exemplu de utilizare Computer Vision

O provocare majoră în viziunea artificială este dezvoltarea unor sisteme care să atingă nivelul perceptual al unui copil mic. Deși tehnologiile actuale pot recunoaște obiecte și urmări mișcarea, inferențele complexe, precum numărarea animalelor sau înțelegerea contextului, rămân dificile. Rețelele neuronale, în special rețelele neuronale convoluționale (CNN), joacă un rol semnificativ în îmbunătățirea acestor procese, permițând sistemelor să învețe automat caracteristici complexe ale imaginilor.

**Clasificarea (Computer Vision)**

Există diverse tehnici de Computer Vision, fiecare fiind adaptată pentru a rezolva sarcini specifice în funcție de aplicații, exemplu de clasificare în Fig.9.   
**3.1 Clasificarea imaginilor**

Una dintre cele mai fundamentale tehnici în viziunea computerizată. Aceasta implică capacitatea unui algoritm de a vedea o imagine și de a o clasifica sau de a prezice cu acuratețe clasa din care face parte imaginea. Algoritmul învață să recunoască tipuri de imagini pe baza unui set de exemple și poate face acest lucru din nou atunci când primește imagini similare.

**3.2 Recunoașterea obiectelor**

Extinde clasificarea imaginilor prin identificarea și localizarea obiectelor specifice dintr-o imagine. Algoritmul detectează clasele de obiecte și le încadrează într-o casetă de delimitare, un dreptunghi care marchează poziția obiectului. Această tehnică este folosită, de exemplu, pentru a găsi și identifica mașini, pietoni sau semne de circulație în imagini.

**3.3 Segmentarea imaginii**

Presupune împărțirea unei imagini în regiuni semnificative sau obiecte distincte, pe baza unor caracteristici comune, precum culoarea, textura, forma sau intensitatea pixelilor. Scopul este de a izola obiectele de interes dintr-o imagine sau dintr-o secvență video, facilitând astfel analiza și înțelegerea mai ușoară a conținutului vizual.

**3.4 Urmărirea obiectelor**

Implică monitorizarea poziției unui obiect în imagini sau videoclipuri după ce acesta a fost detectat. Algoritmul poate prezice și actualiza continuu poziția obiectului în fiecare cadru consecutiv al unui videoclip. Urmărirea obiectelor poate fi aplicată fie pentru un singur obiect, fie pentru mai multe simultan, fiecare fiind urmărit pe o traiectorie distinctă.

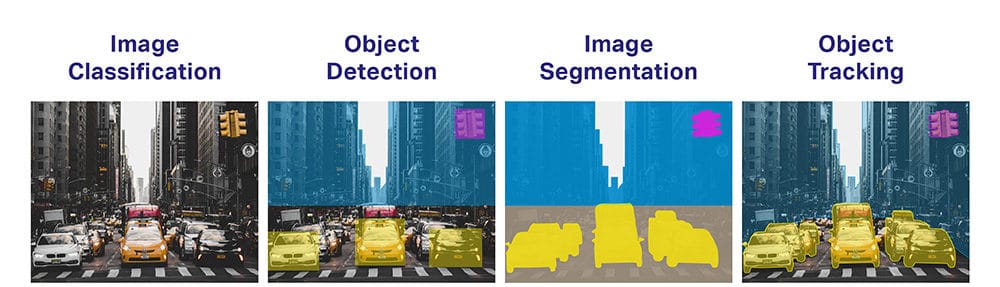
****

Fig.9 Etapele de clasificare a Computer Vision.

**Aplicații ale viziunii computerizate**

Viziunea computerizată evoluează constant, însă multe dintre aplicațiile sale sunt deja larg utilizate în domenii precum afacerile, transportul și sănătatea.Exemple:

**4.1 Google Translate**

Traducerea textul din imagini în timp real. Fixarea camerei telefonului către un semn scris într-o limbă străină, iar aplicația traduce instantaneu textul în limba dorită. Începând cu 2022, Google Translate poate traduce 133 de limbi.

**4.2 Mașini autonome**

Tehnologia dată este esențială pentru funcționarea vehiculelor autonome. Camerele de la bord captează imagini din împrejurimi, iar sistemul procesează aceste date pentru a identifica mașini, pietoni, semne de circulație și alte obiecte. Un exemplu este Autopilotul de la Tesla, care folosește viziunea computerizată pentru a asista conducerea autonomă.

**4.3 Realitatea augmentată (AR)**

Aplicațiile de realitate augmentată care permite detectarea și înțelegerea obiectelor fizice din mediul înconjurător, la integrarea obiectelor generate de computer în lumea reală, în timp ce tehnologia AR oferă o experiență interactivă prin combinarea mediului real cu elemente virtuale, cum ar fi datele GPS, sunete și video.

**4.4 Recunoașterea faciala**

Tehnologia de recunoaștere facială (Fig.9) permite calculatoarelor să asocieze imagini ale oamenilor cu identitățile lor. Aceste aplicații utilizează algoritmi pentru a detecta trăsăturile faciale din imagini. Rețelele neuronale sunt antrenate să recunoască fețele umane și să le distingă de alte obiecte dintr-o imagine. Tehnologia de recunoaștere facială este acum integrată în multe aplicații pe care le folosim zilnic.

****

Fig.10 Face control

**4.5 Supraveghere**

Inteligența artificială și viziunea computerizată au numeroase aplicații în supravegherea video și securitate. Unele dintre aplicațiile de viziune computerizată includ:

* Detectarea oamenilor
* Recunoașterea persoanelor
* Supravegherea vehiculelor
* Detectarea armelor
* Detectarea incidentelor de trafic

Sistemele de supraveghere ajută la creșterea preciziei și la depășirea unor provocări majore, cum ar fi identificarea fețelor sau detectarea unei situații periculoase. De exemplu, prin intermediul camerelor de supraveghere cu rezoluție înaltă, sistemele de supraveghere pot detecta comportamente violente și alerta rapid autoritățile.

**4.6 Sănătate**

Se utilizează în domeniul sănătății pentru diagnosticare. Imagistica prin rezonanță magnetică (IRM) Fig.11, radiografiile și multe alte teste medicale se bazează pe procesarea imaginilor. Algoritmii ajută la procesarea acestor imagini și la detectarea bolilor.



Fig.11 IRM a Emisferei Cerebrale

Aplicațiile de viziune computerizată sunt deja prezente: mașini autonome, realitatea augmentată, recunoașterea facială, supravegherea și sănătatea. Pe măsură ce tehnologia evoluează, din ce în ce mai multe aplicații utilizate pentru a îmbunătăți procesele de afaceri, pentru a crește precizia și pentru a aduce o serie de alte beneficii.

**Analiză comparativă între ochiul uman și o cameră de luat vederi**

**5.1 Structura și mecanismul de captare a imaginii**

|  |  |
| --- | --- |
| **Ochiul uman** | **Camera de luat vederi** |
| Corneea și cristalinul focalizează lumina pe retină, care este formată din celule fotoreceptoare (conuri și bastonașe) care convertesc lumina în semnale electrice | Lentila camerei focalizează lumina pe senzorul digital (CCD sau CMOS), care captează imaginea |

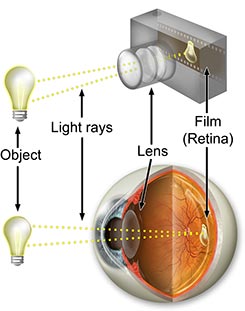


Fig.12 Compararea vizuală a ochiului cu camera

|  |  |
| --- | --- |
| Pupila controlează cantitatea de lumină care intră în ochi, mărindu-se sau micșorându-se în funcție de condițiile de iluminare. | Diafragma camerei controlează cantitatea de lumină care intră, similar cu pupila din ochiul uman. |
| Retina funcționează ca un senzor curbat care transformă informațiile optice în semnale transmise către creier prin nervul optic. | Senzorul este o suprafață plată, și la fel ca retina, convertește lumina în semnale electrice, care sunt apoi procesate și transformate într-o imagine digitală. |
| Ochiul uman percepe culorile prin celulele conuri, care sunt sensibile la trei lungimi de undă principale: roșu, verde și albastru. Această sensibilitate este ajustată de creier, care interpretează variațiile de lumină și culoare într-un mod complex. | Camerele folosesc filtre de culoare, de obicei bazate pe modelul RGB (Red, Green, Blue), pentru a capta culori. Senzorii digitali nu sunt la fel de sensibili la nuanțele subtile și au nevoie de post-procesare pentru a reda culori fidele |

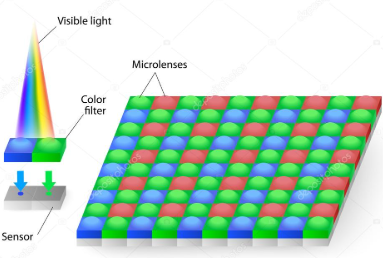
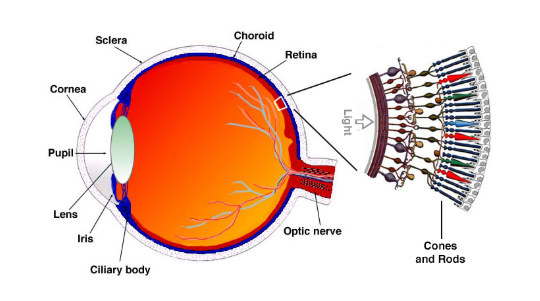
****

Fig.13 Perceperea culorii de către ochiul uman prin celulele conuri și a camerei prin filtre RGB

|  |  |
| --- | --- |
| Ochiul are o capacitate de ajustare naturală la variațiile de lumină, fiind capabil să distingă o gamă largă de culori și nuanțe în diverse condiții de iluminare. | În funcție de calitatea camerei, redarea culorilor poate varia, iar camerele mai simple au o gama dinamică mai restrânsă față de ochiul uman, Fig.13. |
| Ochiul își schimbă focalizarea ajustând forma cristalinului prin procesul de acomodare. Acesta poate trece rapid de la focalizarea pe obiecte apropiate la obiecte îndepărtate. | Cameră focalizează folosind mișcarea lentilei în fața senzorului. În camerele moderne, focalizarea automată se face foarte rapid și precis, dar are nevoie de algoritmi pentru a detecta corect punctul de interes. |

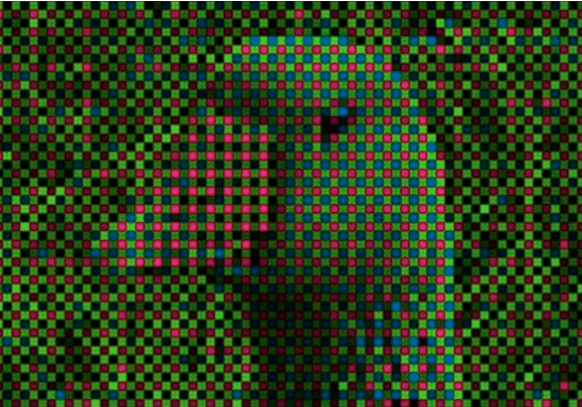
****

Fig.14 Ajustarea imaginii in functie de capacitatile senzorului

Camera de luat vederi și ochiul uman relevă atât asemănări, cât și diferențe semnificative. Ambele captează imagini și permit perceperea mediului înconjurător, dar ochiul uman este mult mai sofisticat în procesarea și interpretarea informațiilor vizuale, fiind capabil să perceapă o gamă mai largă de culori și detalii, precum și să se adapteze rapid la schimbările de lumină.

**Modul de achiziție (Tipuri de senzori)**

În sistemele de luat vederi artificiale (cum ar fi camerele video, dronele sau vehiculele autonome), sunt utilizați diferiți senzori pentru captarea și prelucrarea imaginilor. Cei mai frecvent utilizați senzori sunt:

**6.1 Charged Coupled Device**

Un CCD sau un dispozitiv cuplat cu încărcare este un detector de fotoni foarte sensibil. Este împărțit într-un număr mare de zone mici sensibile la lumină cunoscute sub numele de pixeli, care pot fi folosite pentru a asambla o imagine a zonei de interes. Un CCD este un detector de matrice multicanal pe bază de siliciu de lumină UV, vizibilă și aproape de infrastructură, Fig 15.

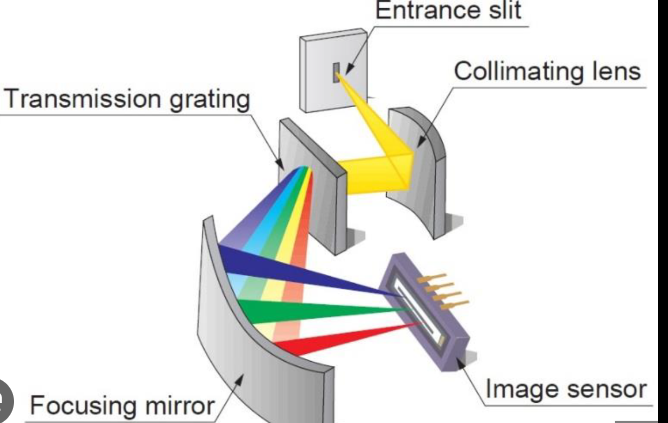


Fig.15 Modul de achizitie a CCD sensor

**6.2 Complementary Metal-Oxide-Semiconductor**

CMOS este un semiconductor complementar de metal-oxid. CMOS este un circuit integrat construit pe o placă de circuit imprimat. Este un cip de memorie alimentat de baterii care păstrează fără efort datele de inițializare. BIOS-ul folosește aceste date pentru a porni dispozitivul, adică în timpul procesului de pornire, Fig 16.

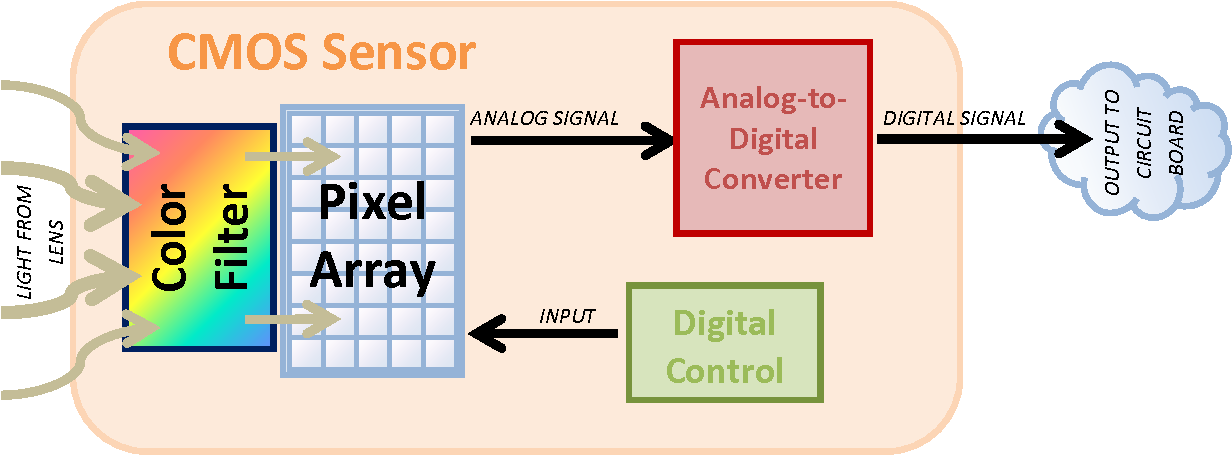


Fig.16 Modul de achizitie a CMOS sensor

**6.3 Light Detection and Ranging**

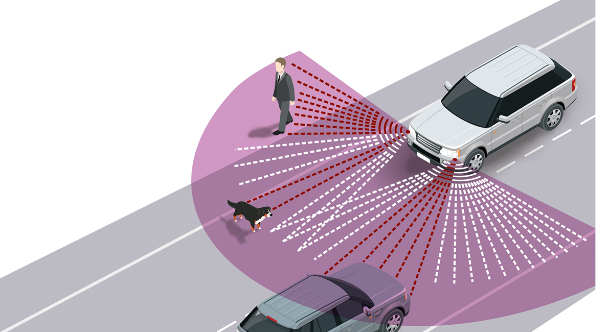
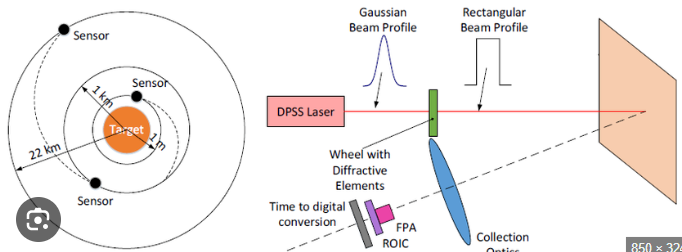
LiDAR este un acronim pentru Light Detection and Ranging. În LiDAR, lumina laser este trimisă de la o sursă (transmițător) și reflectată de obiectele din scenă. Lumina reflectată este detectată de receptorul sistemului și timpul de zbor (TOF) este utilizat pentru a dezvolta o hartă a distanței a obiectelor din scenă, Fig 17.

Fig.17 Modul de achizitie a LiDAR sensor

**6.4 Depth Sensors**

Senzorii de profunzime sunt dispozitive care măsoară distanța dintre senzor și obiectele din mediul înconjurător, oferind informații tridimensionale (3D) despre scenă. Acești senzori nu doar captează imaginea bidimensională a unui obiect, ci și adâncimea (profunzimea), ceea ce permite crearea unui model 3D al mediului, Fig 18.

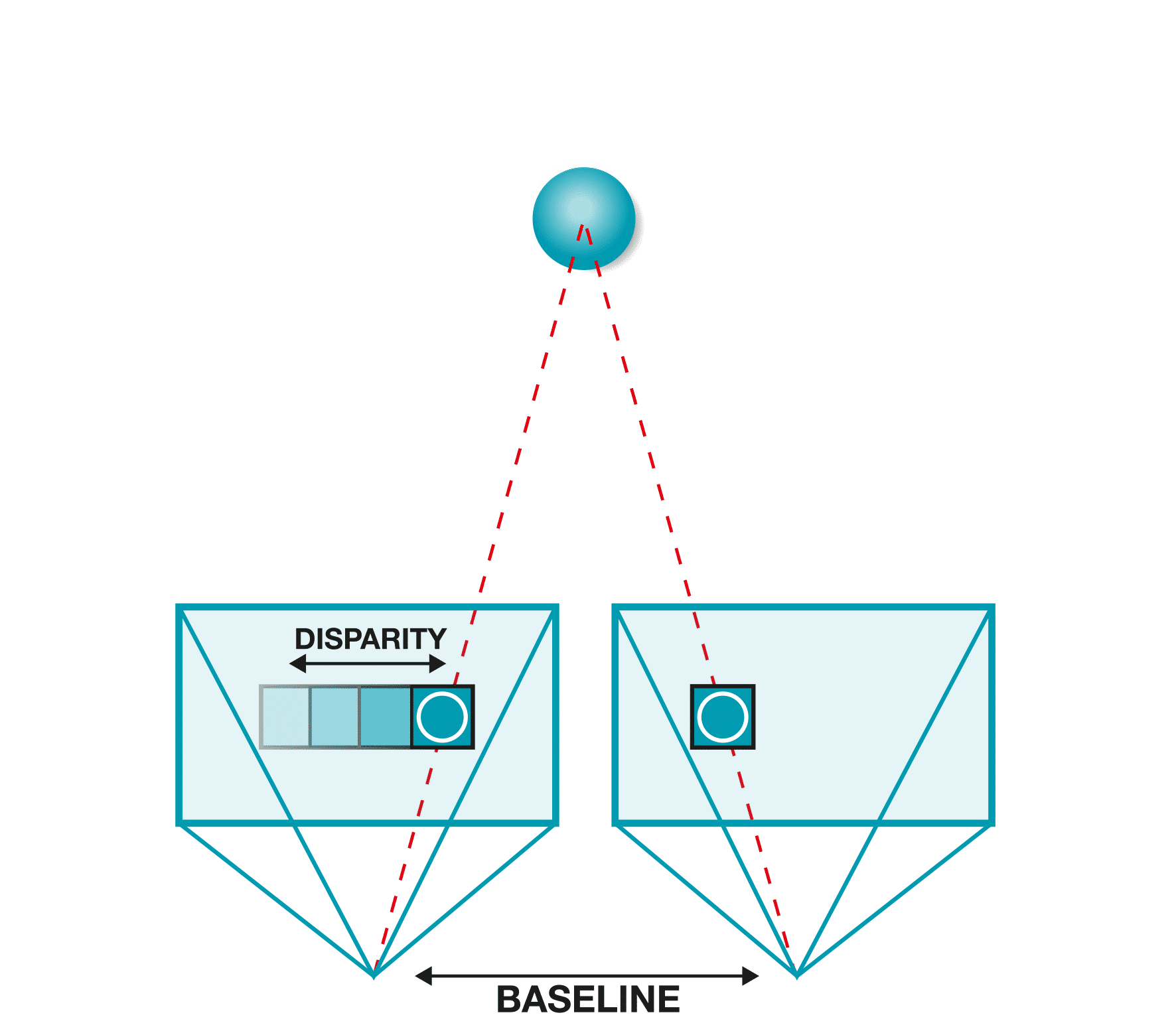
****

Fig.18 Modul de achizitie a Depth Sensors

**Concluzia**

În concluzie, sistemele de vedere artificială au devenit un instrument indispensabil în numeroase domenii, de la securitate și medicină, până la vehicule autonome și realitate augmentată. Prin utilizarea algoritmilor avansați și a tehnologiilor de învățare profundă, aceste sisteme pot analiza, interpreta și procesa date vizuale cu o precizie tot mai mare. Totuși, provocările legate de complexitatea contextelor vizuale și de atingerea nivelului perceptual uman rămân obiective viitoare.

Dezvoltarea continuă în acest domeniu promite să deschidă noi oportunități, contribuind la crearea de soluții mai eficiente și mai inteligente pentru problemele cotidiene și industriale.

**Bibliografie**

"Computer Vision: Algorithms and Applications" – Richard Szeliski

“Digital Image Processing" de Rafael C. Gonzalez și Richard E. Woods

"Pattern Recognition and Machine Learning" – Christopher Bishop

"Learning OpenCV: Computer Vision with the OpenCV Library" de *Gary Bradski și Adrian*