



# Sistem multifocal pentru urmărirea în timp real a trăsăturilor dinamice faciale și corporale (MULTIFACE)

**Raport științific, Etapele 1 și 2.**

**Anul:** 2015 - 2016

**Director proiect:** Conf. Dr. Ing. Radu Dănescu

**Echipa de cercetare:**

Radu Dănescu, Florin Oniga, Diana Borza, Răzvan Itu, Mircea Paul Mureșan

## Cuprins

1. Realizarile etapei anterioare (etapa 1/2015).....	2
2. Rezumatul etapei 2.....	5
3. Realizarea sistemului senzorial multi-focal (finalizare activitate, realizare sistem).....	5
4. Dezvoltarea metodologiei de calibrare preliminară .....	7
5. Dezvoltarea unei tehnici de calibrare eficiente.....	7
6. Stereoviziune pentru urmărirea feței și a capului .....	8
7. Îmbunătățirea algoritmilor de stereoviziune pentru măsurarea precisă a feței umane .....	11
9. Proiectarea modelelor pentru geometria și poziția capului și pentru trăsăturile faciale.....	17
10. Proiectarea și implementarea algoritmilor de urmărire a geometriei și poziției capului faciale. ....	26
11. Diseminarea rezultatelor preliminare .....	27

## 1. Realizarile etapei anterioare (etapa 1/2015)

În prima etapă a proiectului au fost demarate activitățile pentru: realizarea și calibrarea sistemului multifocal, realizarea sistemului de stereoviziune, și modelarea și urmărirea capului și a trasaturilor faciale. Datorită timpului scurt disponibil pentru această etapă, activitățile esențiale planificate inițial în faza de propunere au fost împărțite în sub-activități, în acest an fiind realizate sub-activitățile preliminare, iar în etapa 2 activitățile au fost finalizate.

În etapa 1 a fost realizat un studiu în vederea realizării sistemului senzorial multifocal, și au fost făcute primele achiziții, conținând componente de mică valoare, urmând ca achiziția componentelor principale să fie făcută în etapa 2. A fost definită de asemenea metodologia de calibrare, și au fost studiate performanțele a multipli algoritmi de stereoviziune. A fost făcut de asemenea un studiu detaliat privind modelarea și urmărirea trasaturilor faciale, definind și potențialele contribuții originale punctuale care pot fi aduse.

### *Realizarea sistemului multifocal:*

Pentru realizarea acestui sistem, au fost studiate mai multe opțiuni pentru camere video, lentile, arhitectura de procesare și mecanismele de control și orientare a camerelor. S-a optat pentru un sistem compus din două camere alb-negru, fixe, pentru stereoviziune, montate împreună cu o a treia cameră, care va fi color, și va avea capacitatea de achiziție rapidă (mai mult de 200 de cadre pe secundă). Au fost studiate următoarele variante:

Tabel 1.1. Variante posibile de camere de mare viteză, cu pret accesibil proiectului

Model camera	Senzor	Rezolutie optima	Cadre/secunda (la rezolutia optima)	Caracteristici
Optomotive Velociraptor EVO	CMOSIS CMV2000	2048x1048	340	Marime senzor 2/3", pixel = 5.5 $\mu\text{m}$ , conexiune prin Ethernet, pret ~2700€
Ximea USB3 Vision	CMOSIS CMV300	640x488	500	Marime senzor 1/3", pixel = 7.4 $\mu\text{m}$ , pret ~933€
Megaspeed MS35K	Mega Speed B/W CMOS	640x480	325	Senzor 2/3", conexiune Ethernet, pixel = 7 $\mu\text{m}$ , pret ~9300€

În limita bugetului din 2015, și a timpului scurt pentru implementare acestei etape, au fost efectuate doar achiziții de mică valoare, urmând ca cele care implică un cost mai mare, și implicit o procedură de durată, să fie efectuate în etapa următoare.

În vederea realizării sistemului, au fost achiziționate următoarele componente:

- Motoare pas cu pas
- Plăci SOC de tip Raspberry PI 2
- Plăci cu microcontroller Arduino Mega 2560
- Drive motor de tip Arduino Motor Shield

- Cabluri, surse de alimentare, alte accesorii pentru montaj.

### *Dezvoltarea metodologiei de calibrare preliminară – studiul soluțiilor de calibrare*

În urma studiului soluțiilor de calibrare existente, s-a propus următoarea strategie de calibrare:

- Calibrarea individuală a parametrilor intrinseci pentru fiecare cameră video în parte, folosind metoda Caltech.
- Calibrarea parametrilor extrinseci ai sistemului stereo, folosind metoda proprie bazată pe repere 3D. Această calibrare va fi repetată doar dacă există indicii temeinice privind degradarea performanței de reconstrucție, sau dacă se modifică poziția camerelor.
- Calibrarea parametrilor extrinseci ai camerei de viteză mare, pentru mai multe poziții ale sistemului pan-tilt, folosind aceeași metodă ca pentru camerele sistemului stereo.
- Calibrarea automată, în timp real, a camerei mobile, pe baza informației 3D furnizată de sistemul stereo, coroborată cu trăsături invariante extrase din cele trei imagini și puse în corespondență (colțuri, trăsături în mișcare, etc).

### *Evaluarea acurateții algoritmilor de stereoviziune*

Mai mulți algoritmi de căutare a corespondențelor stereo dense au fost evaluați folosind bazele de date Middlebury și Kitti. Algoritmii stereo se pot clasifica în algoritmi globali, locali și semi-globali. Algoritmii locali sunt mai rapizi, necesită o cantitate de memorie redusă, dar au performanțe de calitate limitată. Cei globali impun constrângeri globale rezultatelor corespondenței stereo, dar acest lucru vine cu un cost suplimentar. Algoritmii semi-globali combină avantajele și dezavantajele celor două metode, putând fi rulați în timp real pe arhitecturi de tip GPU.

Au fost folosite mai multe metrice de evaluare:

Out-All – Aceasta eroare masoara diferenta dintre un pixel reconstruit si unul de ground truth. Daca eroarea este mai mare de un anumit threshold (2 sau 3 pixeli) pixelul reconstruit va fi considerat eronat. Aceasta metoda este folosita cel mai des.

Density – Acest parametru ne informeaza despre numarul de puncte reconstruite. Adica pentru fiecare punct de ground truth daca avem un punct reconstruit se considera ca avem densitate pentru respectivul punct. Cand se foloseste, aceasta metrica, nu ne intereseaza acuratetea din punctual respective.

Average – Eroarea medie ne spune pentru fiecare pixel reconstruit correct, care este eroarea fata de ground truth.

Nonocc – se testeaza care este eroarea doar in zonele care nu sunt ocluate, adica zonele care se vad din ambele imagini. Figura de mai jos arata intuitive zonele la care se refera aceasta metrica pentru primul set de imagini de mai sus. Cu negru este evidentiata zona in care exista ocluzii.

Disc – se testeaza regiunile in apropierea zonelor cu dicontinuitati.

În urma testelor, s-a demonstrat că algoritmii de tip SGM (semi-global matching) sunt cea mai bună combinație de calitate și cost, dar că metrica pentru similaritate influențează foarte mult rezultatele.

Din rezultatul analizei facute am dedus ca nu exista o metoda “cea mai buna” care sa rezolve toate problemele. Dupa realizarea sistemului senzorial multifocal, se vor experimenta mai multi algoritmi si se va alege cel care ofera cel mai bun raport dintre acuratete si timp de executie.

*Proiectarea modelelor pentru geometria și poziția capului, și pentru trăsăturile faciale:  
Studiul modelelor și a tehnicilor de urmărire existente pentru cap și trăsături faciale*

Au fost studiate mai multe metode pentru reprezentarea expresiilor și a microexpresiilor faciale, iar cel mai reprezentativ model a fost determinat a fi modelul Candide 3, care poate reprezenta majoritatea fetelor, și care poate fi reglat pentru înălțimea capului, poziția sprâncenelor, poziția verticală a nasului. De asemenea, modelul poate include unități de animație (AU), care sunt deformări ale formei neutre. Modelul Candide este des folosit în domeniul viziunii artificiale, datorită simplității sale, care permite procesarea și recunoașterea expresiilor faciale cu o complexitate redusă. De exemplu, Face Tracking SDK al senzorului Microsoft Kinect utilizează modelul Candide 3 pentru a urmări fața oamenilor în timp real.

Au fost definite următoarele contribuții originale ce pot fi aduse:

- Sistemul dezvoltat nu va fi amplasat pe capul subiectului (nu va fi head-mounted), fiind un sistem non-invaziv. Astfel, subiectul nu va fi conștient de faptul că este analizat și nici nu va resimți vreun inconvenient în acest proces. Sistemele head-mounted sunt dificil de utilizat mai ales în cazul copiilor, care au tendința de a le îndepărta.

- Nu se vor utiliza surse de lumină în infraroșu (IR). Majoritatea sistemelor de urmărire ale ochilor se bazează pe surse de lumină IR, deoarece acestea nu sunt vizibile ochiului uman și facilitează procesul de detecție. Principalul dezavantaj al acestor metode este faptul că nu pot fi utilizate în mediul exterior, datorită interferențelor legate de iluminare.

- Rezultatele acestui modul vor fi corelate cu alte informații legate de orientarea capului, prezența unor ocluziuni, poziția corpului, expresii faciale etc. Astfel, sistemul dezvoltat va fi mult mai tolerant față de ocluziuni, cum ar fi ochelarii sau condițiile de iluminare. În plus, prin combinarea acestor informații se va obține o mai bună înțelegere a proceselor emoționale și cognitive ale oamenilor.

- Rezultatele obținute vor fi analizate pentru a determina diferiți parametri: dimensiunea pupilei în diferite momente de timp, direcția de privire, frecvența clipirilor, distanța interpupilară etc. Acești parametri au o importanță deosebită pentru o mulțime de aplicații cum ar fi: optometria, determinarea gradului de dificultate cognitivă al unei sarcini (cognitive workload), detecția oboselii etc.

- Folosind o cameră high-speed se vor detecta și se vor determina parametri sacadelor: amplitudinea, viteza maximă (peak velocity), durata. Analiza acestor mișcări are un impact important atât în medicină, cât și în înțelegerea anumitor procese psihologice. Mai multe studii medicale au arătat că anumite diferențe în mișcările ochilor pot indica prezența anumitor tulburări de neuro-dezvoltare, cum ar fi: deficitul de atenție, autismul, sindromul alcoolismului fetal sau al unor boli neurologice.

### *Resursa umană*

În vederea constituirii echipei independente de cercetare, în această etapă au fost angajați doi cercetători doctoranzi care nu au mai avut contract de muncă cu Universitatea Tehnică, și un cercetător masterand.

Conform planului de realizare, activitatea de cercetare desfășurată în această etapă răspunde cerințelor proiectului, și considerăm ca obiectivele planificate pentru acest an sunt îndeplinite.

## 2. Rezumatul etapei 2

În această etapă au fost continuate activitățile începute în anul anterior în vederea realizării obiectivelor proiectului: finalizarea realizării sistemului senzorial multifocal, dezvoltarea unei metodologii eficiente de calibrare (cu intervenție minimă din partea utilizatorului), adaptarea algoritmilor de stereoviziune pentru măsurarea precisă a feței umane, modelarea și urmărirea poziției capului (folosind informația stereo) și detecția și urmărirea trăsăturilor faciale.

Pentru îmbunătățirea calității reconstrucției tridimensionale, au fost proiectate și implementate metode noi pentru calculul corespondențelor stereo, care au reușit să producă rezultate cu densitate superioară, fără a folosi însă metode de optimizare globală.

În această etapă s-a realizat achiziția principală a acestui proiect: două camere de viteză mare (*high speed*) XIMEA. În plus s-au mai achiziționat anumite componente mecanice necesare realizării prototipului unității *pan tilt*.

S-a stabilit un model 3D pentru fața umană care va fi utilizat pentru a urmări trăsăturile faciale și pentru a calcula orientarea capului. Pentru înțelegerea modului în care ar trebui potrivit modelul 3D peste imagini, s-a dezvoltat o aplicație care permite suprapunerea manuală a modelului peste o imagine. Pe baza observațiilor obținute, s-a proiectat și s-a implementat un algoritm care potrivește automat și rapid modelul peste o imagine, determinându-se astfel orientarea capului și localizarea elementelor feței (ochi, buze, sprâncene).

De asemenea, am dezvoltat 2 metode originale de detecție a trăsăturilor ochilor (centrul irisului, colțurile ochilor și pleoapele) în imagini și o metodă originală de urmărire a acestor trăsături și de detecție a clipirilor în cadre video.

Activitatea de cercetare pe care am desfășurat-o în această etapă corespunde planului de realizare, iar cele mai importante rezultate obținute au fost publicate la conferințe și reviste de specialitate. În concluzie, considerăm că obiectivele principale planificate în acest an au fost îndeplinite.

## 3. Realizarea sistemului senzorial multi-focal (finalizare activitate, realizare sistem)

Unul din obiectivele esențiale ale proiectului a fost realizarea sistemului senzorial multifocal propus. În acest sens s-a proiectat un suport (*rig*) cu distanța dintre camere (*baseline*) variabilă pentru susținerea celor două camere care compun sistemul stereo. O imagine a rig-ului este ilustrată în Figura 1.

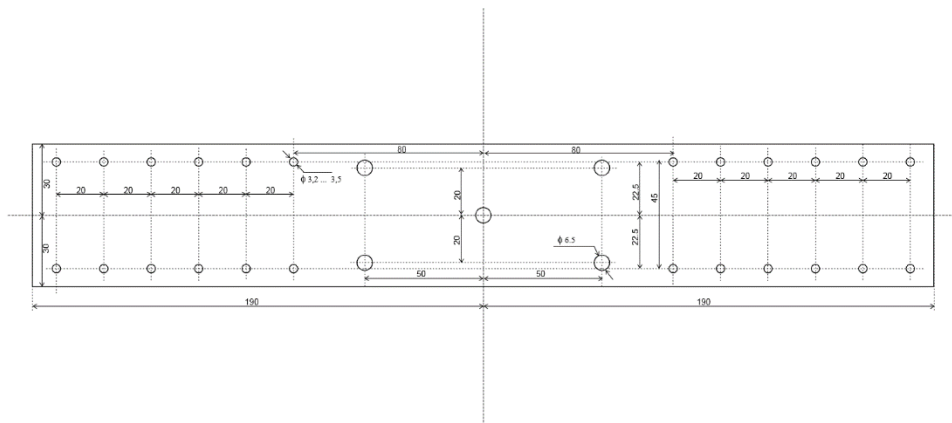


Figura 1. Rig rigid pentru susținerea sistemului de stereo-viziune

Pentru montarea camerei high-speed a fost creat încă un sistem mecanic auxiliar care se fixează în rigul stereo. Sistemul multifocal final se poate observa în Figura 2. Cele două camere montate în paralel una față de cealaltă vor avea ca scop reconstrucția 3D a scenei, iar a treia cameră are ca scop detecția și urmărirea cu viteză mare a fețelor și a trăsăturilor faciale folosind informațiile obținute de la sistemul stereo.

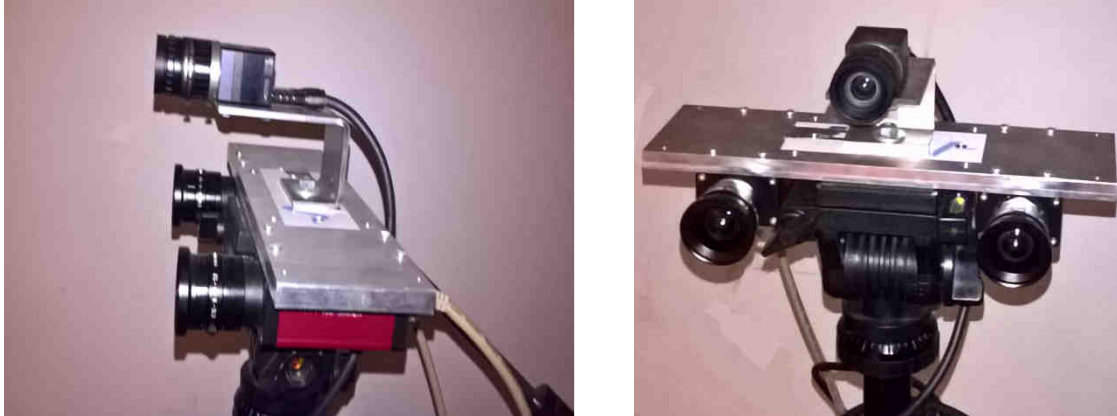
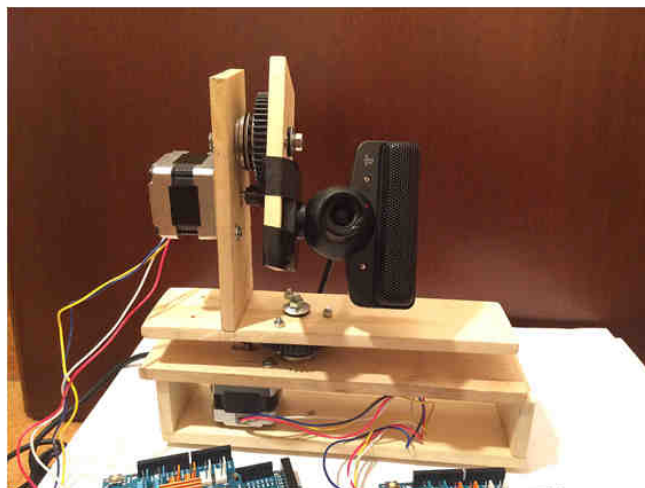


Figura 2. Sistemul trifocal

Arhitectura software realizată, prezintă atât funcții de nivel înalt cât și funcții de nivel jos. Funcțiile de nivel jos cuprind: achiziție, calibrare, rectificare, reconstrucție 3D, flux optic. Funcțiile de nivel înalt cuprind: recunoașterea obstacolelor, extragerea trăsăturilor, detecția fețelor, urmărirea fețelor, extragerea micro-expresiilor.

Achiziția imaginilor de la fiecare cameră a fost implementată pentru platforme care funcționează pe 32 și 64 de biți, iar camerele au fost sincronizate cu ajutorul unui impuls software folosind protocolul PTP (Precision Time Protocol).

De asemenea, am dezvoltat și un prototip pentru unitatea *pan tilt*, care va orienta a treia cameră astfel încât aceasta să perceapă doar zona capului. Acest prototip este ilustrat în Figura 3. Sistemul are două grade de libertate (două unghiuri de rotație, în plan orizontal și vertical), poziționarea precisă fiind obținută folosind motoare pas cu pas controlate de microcontrollere.



#### 4. Dezvoltarea metodologiei de calibrare preliminară

Prin procesul de calibrare se determină atât parametrii intrinseci cât și cei extrinseci ai fiecărei camere. Acest proces de calibrare este realizat o singură dată pentru fiecare cameră. Procesul de calibrare se realizează cu ajutorul uneltelor de calibrare Caltech. Potrivit indicațiilor acestora, pentru a extrage parametrii intrinseci se utilizează imagini ale unui pattern de calibrare - similar cu o tabla de șah - capturate dintr-o serie de unghiuri (perspective). Fundalul și placa trebuie să fie mate și rigide în așa fel încât să nu existe reflexii și imaginea capturată să nu fie saturată.

Cea mai bună metodologie este ca paternul să fie așezat pe arcuri de cerc la distanțe variabile astfel încât estimarea distanței focale să fie cât mai precisă. Distanța la care este plasat paternul trebuie limitată astfel încât suprafața pattern-ului în imagine să nu fie mai mică de 1/6 din dimensiunea imaginii. Cu cât distanța la care este plasat pattern-ul va crește cu atât eroarea de măsurare a punctelor de control din imagine va crește. Distanța minimă este limitată de distorsiunile introduse de camera și lentile. În Figura 4 se observă pozițiile la care trebuie așezat paternul de calibrare cât și un exemplu de pattern. Pentru extragerea parametrilor extrinseci se plasează la distanțe cunoscute un număr de ținte și este setat un centru de coordonate al sistemului care este plasat în mijlocul sistemului stereo. Folosind parametrii intrinseci obținuți, imaginea cu ținte și cunoscând distanțele la fiecare țintă se determină parametrii extrinseci ai sistemului trifocal. Pentru determinarea parametrilor extrinseci se folosește același instrument de calibrare oferit de Caltech.

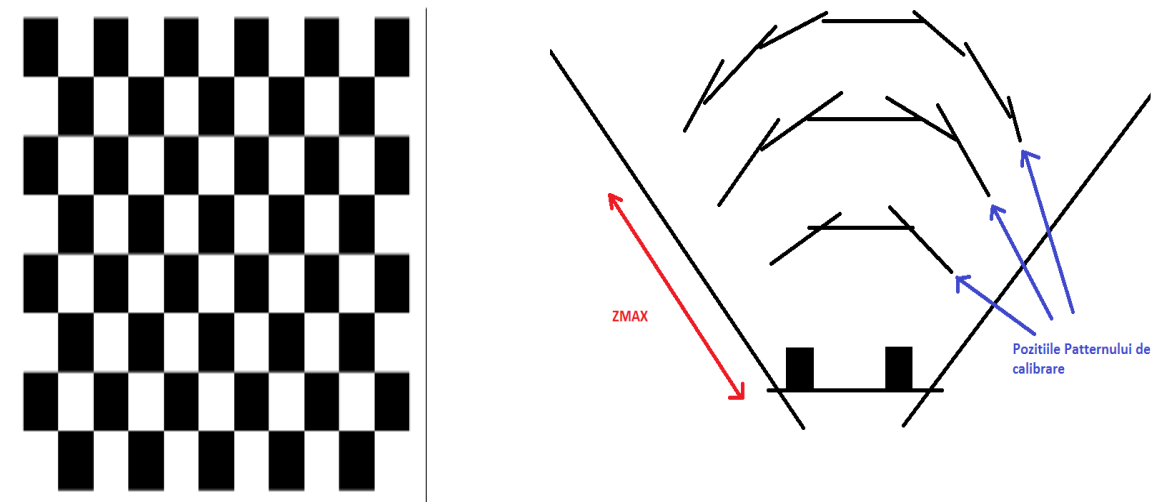


Figura 4. Pattern de calibrare și pozițiile paternului de calibrare

#### 5. Dezvoltarea unei tehnici de calibrare eficiente

Pentru dezvoltarea unei tehnici de calibrare care să elimine, pe cât posibil, intervenția umană, vom porni de la presupunerea că parametrii intrinseci ai camerelor sunt calibrați. Acest pas se execută o singură dată pentru o cameră, nefiind necesar să se repete atunci când geometria sistemului este schimbată. În consecință, vom defini problema calibrării automate ca problema calibrării automate a parametrilor extrinseci, matricea de rotație și vectorul de translație al fiecărei camere a sistemului multifocal.

Pentru determinarea automată a parametrilor extrinseci, este nevoie ca un o scenă țintă să fie identificată automat în toate cele trei camere ale sistemului. De asemenea, scena trebuie să prezinte

suficiente detalii pentru a putea infera parametrii. O astfel de scenă, cu o complexitate ridicată și cu posibilitatea potrivirii automate pe imagini, este fața umană, al cărei model parametric este descris în secțiunea 8 a acestui raport, și care se poate potrivi automat pe imagine folosind metoda descrisă în figura 9. Deoarece potrivirea implică automat găsirea parametrilor intrinseci pentru procesul de proiectie, o potrivire cu succes pe toate cele trei imagini va duce la găsirea parametrilor extrinseci relativi între cele trei camere.

## 6. Stereoviziune pentru urmărirea feței și a capului

Procesul de rectificare e important întrucât reduce spațiul de căutare folosit în corespondența stereo. Rectificarea face ca imaginile achiziționate de un sistem stereo să pară că ar fi luate cu aceeași set de camere care au parametrii identici. De asemenea, prin procesul de rectificare, liniile epipolare vor fi transformate în așa fel încât să se suprapună cu liniile imaginii. Procesul de stereo-corespondența se simplifică mult, întrucât căutarea trăsăturilor similare în cele două imagini este efectuată pe liniile imaginii (tipul de căutare pentru o trăsătură devenind liniar). Ideea de bază a rectificării este de a defini un nou sistem care este canonic și de a calcula noii parametri având în vedere constrângerile impuse de un astfel de sistem.

Constrângerile sistemului canonic sunt următoarele:

1. Planurile imaginii trebuie să fie coplanare
2. Axe optice trebuie să fie paralele
3. Axele  $x$  ale noului sistem (al imaginilor rectificate) trebuie să fie coliniare.
4. Distanțele focale trebuie să fie egale
5. Coordonate punctului principal trebuie să fie identic în cele două imagini
6. Centrele optice ale sistemului rectificat trebuie să rămână neschimbate
7. Proiecția fiecărui punct trebuie să aibă aceeași coordonată verticală

După montarea rigului stereo trebuie definite axele și orientarea sistemului de coordonate global (*world coordinate system*). Convenția utilizată este ilustrată în Figura 5.

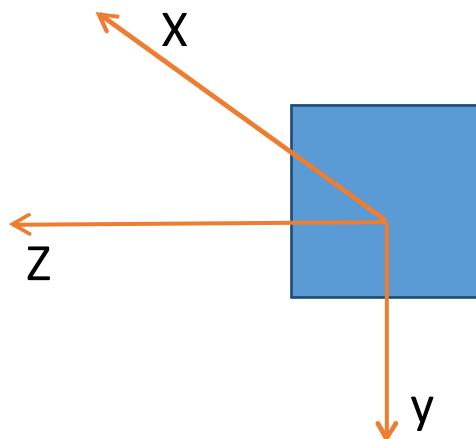


Figura 5. Convenția adoptată pentru orientarea axelor în sistemul de coordonate global și camerelor

Parametrii noului sistem canonic sunt calculați folosindu-ne de vechii parametri având în vedere orientarea axelor. Odată ce parametrii sistemului canonic au fost calculați, imaginile originale achiziționate cu sistemul stereo (imaginile nerectificate), trebuie să fie transformate astfel încât să pară că au fost achiziționate cu noul sistem. Această transformare



este realizată reprojectând fiecare punct din imaginile originale în sistemul lumii, iar apoi proiectând-le înapoi în imaginile rectificcate. Aceasta idee de bază este ilustrată în Figura 6.

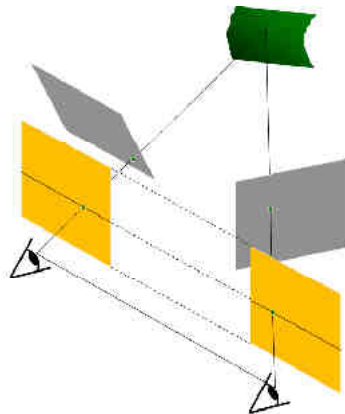


Figura 6. Procesul de rectificcare explicat în mod intuitiv

Procesul de rectificcare poate fi împărțit intuitiv în două etape. În prima etapă, etapa offline, se calculează parametrii noului sistem canonic. În a doua etapă, etapa online, fiecare imagine trebuie transformată.

#### I. Etapa offline

În primul rând având în vedere constrângerea numărul 4 putem calcula noua distanță focală făcând media aritmetică a distanțelor focale originale.

$$fxCan = \frac{flx + frx}{2}$$

$$fyCan = \frac{fly + fry}{2}$$

În al doilea rând, având în vedere constrângerea 5 putem calcula noul punct principal ca fiind centrul imaginii sau să fie media aritmetică a coordonatelor punctelor principale inițiale.

$$oxCan = \frac{olx + orx}{2}$$

$$oyCan = \frac{oly + ory}{2}$$

Având în vedere parametrii canonici, se poate calcula noua matrice internă canonică.

$$ACan = \begin{matrix} & \begin{matrix} fxCan & y & oxCan \end{matrix} \\ \begin{matrix} 0 & fyCan & oyCan \\ 0 & 0 & 1 \end{matrix} & \end{matrix}$$

Al treilea pas constă în calcularea noilor axe de coordonate. Considerând constrângerile 3 și 6 noua axă x poate fi considerată linia de bază a sistemului vechi ( linia determinată de cei doi centrii optici ). Noua axa oy este calculată din produsul vectorial:

$$\vec{oy} = \vec{ozOld} \times \vec{ox}$$

Pentru a calcula noua axă y se alege una din axele z ale vechiului sistem iar apoi se calculează noua axă y ca fiind axa care este ortogonală pe planul format de axa oz și ox.

În implementarea noastră am ales axa oz a camerei stângi. Noua axă z este calculată ca fiind axa care este ortogonală pe planul format din noua axă oy și noua axă ox.

Fiecare din cele 3 axe este normalizată cu norma vectorului pe care o reprezintă. Al patrulea pas constă în calcularea noii matrice de rotație. Matricea canonică de rotație este compusă din cei trei vectori calculați precedent. Matricele de transformare sunt calculate după următoarele formule :

$$T_{left} = A_{can} * R_{wcCan} * R_{wcLeft}^T * A_{ol}^{-1}$$

$$T_{right} = A_{can} * R_{wcCan} * R_{wcRight}^T * A_{or}^{-1}$$

Matricea de rectificare are rolul de a lua un punct din imagine, de a calcula coordonatele lui normalizate în sistemul de coordonate global și o reproiectează în imaginea rectificată. Fără a lua în considerare translația dintre sistemele de referință a lumii și a camerei întrucât centrele optice nu se schimbă. Înmulțirea cu  $A^{-1}$  este folosită pentru a transforma punctul din imagine în coordonate normalizate în sistemul de referință al camerei. Pentru rotirea punctului în coordonatele lumii normalizate se execută înmulțirea cu  $R_{wc}$ . Pentru a roti punctul în sistemul canonic de referință se execută înmulțirea cu  $R_{wcCan}$ . În final, pentru a reproiecta punctul în planul imaginii rectificate se înmulțește cu  $A_{can}$ . Înainte de reproiectarea punctului în planul imaginii se aplica operații de eliminare a distorsiunilor (caț timp punctul e în planul lumii). În acest sens se aplică operații de reducere a distorsiunii tangențiale și radiale. Ultimul pas al rectificării ar fi calcularea poziției noi a fiecărui pixel folosind matricea de transformare. Acest lucru este realizat cu un set de look-up table-uri pentru a face transformarea mai rapidă. Rezultatul final este calculat prin interpolare bilineară.

## II. Etapa online

În etapa online intensitatea pixelilor din imaginea rectificată se calculează folosind interpolare bilineara. Acest lucru este realizat întrucât este posibil ca anumiți pixeli din imaginea rectificată sa nu aibă corespondent în imaginea originală (Figura 7).

Interpolarea bilineara este realizată conform formulelor de mai jos:

$$f(x,y) = f(0,0)(1-x)(1-y) + f(1,0)x(1-y) + f(0,1)(1-x)y + f(1,1)xy$$

Imaginile rectificate pot fi observate în Figura 7. Fiecare trăsătură este la aceleași coordonate pe axa y.



Figura 7. Imagini rectificate

Un alt aspect realizat în sensul procesării mai eficiente îl reprezintă scalarea imaginii. Acest coeficient de scalare este introdus automat în matricea internă a imaginii originale. Aducând imaginea mai aproape de centrul optic și modificând coordonatele punctului principal luând în considerare factorul de scalare, imaginile noi rectificatice sunt și redimensionate. O imagine intuitivă a procesului de scalare îl întâlnim în Figura 8.

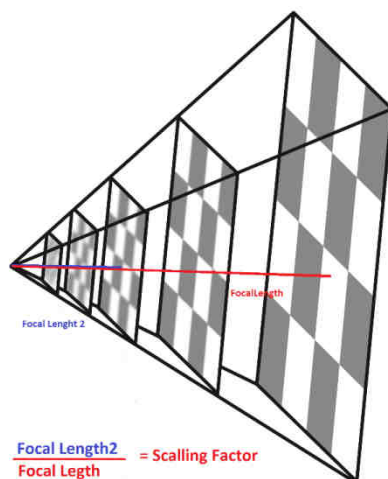


Figura 8. Modul de scalare al imaginilor

Multiplicarea distanței focale și a punctului principal cu factorul de scalare și integrarea noilor valori în matricea internă sunt singurele modificări de care este nevoie pentru a obține o procedură de scalare foarte eficientă. Aceste modificări se realizează în etapa offline.

Corelarea stereo densă reprezintă o funcționalitate de nivel jos, care este folosită ca intrare pentru funcțiile de nivel înalt și care ajută la reconstrucția 3D a mediului. Cele două camere care sunt poziționate la o distanță una față de cealaltă achiziționează imagini în mod sincron. Algoritmul de corelare caută pentru fiecare pixel din o imaginea stângă pixelul corespondent din imaginea dreaptă. Pentru eficientizarea căutării, corespondența nu se face pixel la pixel, ci în jurul fiecărui pixel se alege o regiune de interes și aceasta se folosește pentru a căuta corespondențele. Pentru eliminarea zgomotului, se folosește o agregare rectangulară iar apoi, pentru rafinare de regulă se folosește algoritmul *SGM* care netezește imaginea de disparitate.

Pentru a face căutarea corespondențelor mai eficientă, mai portabilă și cu un consum de resurse hardware mai mic s-a dezvoltat o metodă originală de corelare stereo.

## 7. Îmbunătățirea algoritmilor de stereoviziune pentru măsurarea precisă a feței umane

Metodele locale de stereo corespondență funcționează prin crearea unor regiuni de suport finite, numite blocuri, în jurul fiecărui pixel. După crearea blocurilor pentru imaginile stângă și dreaptă, se caută corespondențele pe liniile epipolare în ambele imagini. Aceste metode se bazează pe o metrică de potrivire și aplică diverse metode de agregare pentru netezire. Cele mai comune metrici de potrivire sunt: *sum of absolute differences* (SAD), *sum of squared differences* (SSD), *rank transform* (RT), *census transform* (CT) sau *normalized cross*

*correlation* (NCC). Dezavantajul metodelor locale față de alte tipuri de metode (globale sau semi-globale) constă în densitate și precizie. Cu toate acestea algoritmi locali consumă resurse puține și nu necesită metode de accelerare hardware pentru a putea funcționa în timp real, ei putând funcționa în timp real și pe CPU.

Unul din motivele pentru care algoritmi de *block matching* nu reușesc să reconstruiască anumite suprafețe este din cauza presupunerii că distanța 3D în toate punctele unui bloc de potrivire e aceeași. În acest sens s-a dezvoltat o metodă care reduce erorile cauzate de orientările diferite ale suprafețelor. Aceasta operație este efectuată foarte rapid folosind un set de look-up table. După calcularea costului de potrivire, rezultatele obținute sunt agregate folosind blocuri de lungimi și forme diferite. Pentru a asigura o calitate superioară a rezultatelor s-au introdus mai multe constrângeri care au ca scop eliminarea de la început a disparităților eronate. Pentru a surprinde înclinarea suprafețelor s-au folosit șapte look-up table-uri care conțin derivata poziției în raport cu un pixel central. Pe baza acestor tabele se construiesc șapte imagini descriptor (o imagine care corespunde cazului fronto-paralel iar șase corespund cazului înclinat) pentru imaginea de intensitate dreaptă. Pentru imaginea de intensitate stângă se creează doar o singură imagine descriptor și anume pentru cazul fronto-paralel. Pentru algoritmul nostru am folosit descriptori binari întrucât aceștia sunt invarianți la deplasamente aditive și multiplicative în intensitate. De asemenea, s-a proiectat un descriptor simetric față de centru care asigură o robustețe crescută în special în cazul pixelilor de zgomot. Expresia descriptorului este ilustrată în ecuația de mai jos:

$$MCST\_CENSUS(u, v) = \sum_{i=0}^{\frac{N-1}{2}} \sum_{j=0}^M \varepsilon(I(u_i, v_j), I(u_{N-i}, v_{M-j-1})) \oplus \sum_{j=0}^{M/2} \varepsilon(I(u_{N/2}, v_j), I(u_{N/2}, v_{M-j-1}))$$

$$\varepsilon(X, Y) = \begin{cases} 00, X - t \geq Y \\ 01, X - t < Y \text{ AND } X + t \geq Y \\ 11, X + t < Y \end{cases}$$

Variabila  $t$  reprezintă un prag mic de intensitate cu ajutorul căruia compensăm variații mici în intensitate. În soluția noastră valoarea lui  $t$  a fost setată la 1. O imagine intuitivă cu lookup table-urile este ilustrată în Figura 9.

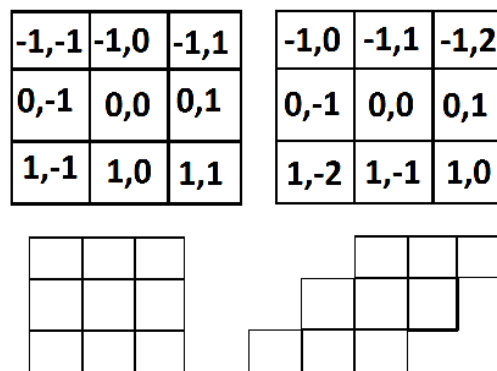


Figura 9. Imagine intuitivă a unui look-up table pentru calculul descriptorului

Pentru calculul costului de corelare între imaginea descriptor stângă și dreaptă s-a creat o metodă de calcul bazată pe distanța Hamming. În ecuațiile de mai jos se reprezintă scorul minim de corelare între imaginea stângă și cele șase imagini descriptor inclinate (SRI). Prin

variabila *Frontal* reprezentam scorul obținut pentru corelarea fronto-paralela a imaginilor descriptor stânga-dreapta (*FPRI* reprezintă imaginea descriptor dreapta pentru cazul fronto-paralel). Rezultatul final al algoritmului nostru pentru un pixel la poziția  $(I,j)$  având disparitatea  $d$  e notat cu  $FS(I,j,d)$  și e obținut prin luarea valorii minime dintre scorul pentru cazul frontal și cel pentru cazul inclinat la care se adaugă o mică penalitate. Penalitatea are rolul de a favoriza suprafețele înclinate.

$$SI(i, j, d) = \sum_{k=1}^6 \text{Min}(\text{Hamming}(\text{IL}(i, j), \text{SRI}_k(i, j, d)))$$

$$\text{Frontal}(i, j, d) = \text{Hamming}(\text{IL}(i, j), \text{FPRI}(i, j, d))$$

$$FS(i, j, d) = \text{Min}(\text{Frontal}(i, j, d), \text{PS}(i, j, d) + \text{Penalty})$$

Pentru filtrarea erorilor se aplică o agregare multi-block folosind blocuri de dimensiune variabilă:

$$A = \max(\text{val}_{1 \times 155}, \text{val}_{155 \times 1}) \text{val}_{17 \times 17} \text{val}_{7 \times 7}$$

Rezultatul final se obține printr-o metoda *greedy* selectând pentru fiecare poziție  $i$  și  $j$  valoarea disparității unde costul este minim. În timpul căutării disparității minime se impun anumite constrângeri de periodicitate și confidență asupra valorilor candidat, care îmbunătățesc calitatea finală a hărții de disparitate. Pentru acuratețe pe lângă metodele clasice de rafinare se mai aplică și o interpolare la nivel de sub-pixel folosind metoda *simetric V*. Expresia analitică a interpolării la nivel de sub-pixel poate fi observată în expresia de mai jos.

$$Disp_{final} = Disp_{integer} + \begin{cases} 0.5 - 0.25 \cdot \left( \frac{(M_3 - M_1)^2}{(M_2 - M_1)^2} + \frac{M_3 - M_1}{M_2 - M_1} \right) & , \text{if } M_2 > M_3 \\ - \left( 0.5 - 0.25 \cdot \left( \frac{(M_2 - M_1)^2}{(M_3 - M_1)^2} + \frac{M_2 - M_1}{M_3 - M_1} \right) \right) & , \text{if } M_2 \leq M_3 \end{cases}$$

În tabelul de mai jos metoda de stereo corespondența propusă (*shifted-block matching - SBM*) este comparată calitativ cu alte metode de stereo corespondență locale care folosesc descriptori binari. Se observă că metoda are densitate ridicată și o eroare mult mai mică în comparație cu ale metode.

Metoda	Rezultate pe imagini din baza de date Kiti	
	Densitatea	Eroarea totală
Census	99.67%	11.865%
WCS-CENSUS	99.98%	11.24%
MCT	99.68%	11.21%
<b>SBM</b>	<b>99.97%</b>	<b>8.785%</b>
Sparse Census	99.51%	13.29%

Exemple de imagini de disparitate sunt ilustrate în Figura 10 și Figura 11.



Figura 10. Imagini de disparitate

Punctele 3D reconstruite au fost proiectate în spațiul lumii pentru a putea fi folosite în procesări ulterioare.

În cadrul aplicației, s-a mai dezvoltat un soft de vizualizare și manipulare a punctelor 3D. Imaginile reconstruite folosind această funcționalitate pot fi observate în Figura 11. Softul de proiecție mai are și rolul de a ajuta la evaluarea metodelor de reconstruite.

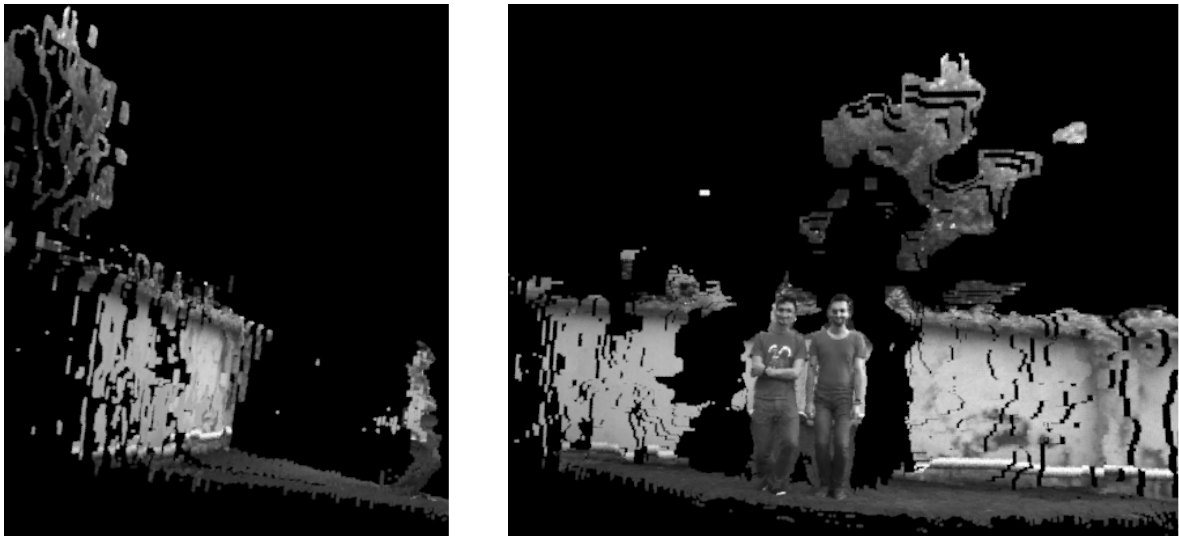


Figura 11. Vedere din lateral și din fata a punctelor 3D reconstruite

## **8. Modelarea și urmărirea geometriei și a poziției capului folosind informația stereo**

Folosind informația obținută prin corespondența stereo putem găsi obiectele din imagine și poziția acestora. După găsirea obiectelor din imagine informațiile obținute vor fi proiectate în camera a treia unde vor avea ajuta la detecția mai rapidă și mai robustă a fețelor.

Pentru detecția obstacolelor s-a implementat o soluție bazată pe algoritmul de disparitate UV (*UV disparity*). *U disparity* reprezintă un acumulator ce conține numărul de voturi asociat fiecărei disparități de pe coloana din matricea de disparitate, în timp ce *V disparity* reprezintă un acumulator ce conține numărul de voturi asociat fiecărei disparități de pe linia din matricea de disparitate.

Acumulatorul *V disparity* va fi utilizat pentru extragerea detaliilor despre suprafața drumului, sau a suprafeței podelei în cazul scenelor de interior. Acest lucru este important întrucât ajută la detectarea planului pe care vor evolua persoanele detectate. De asemenea, folosind acumulatorul *V* putem filtra rezultatele obținute, lăsând doar acele obiecte care au contact cu drumul/podeaua. Pentru calculul acumulatorului se va parcurge vectorul de puncte ce conține informații cu privire la disparitate și se incrementează poziția corespunzătoare în matrice.

$$V_{jd} = \sum_{I=0}^{COLS} \Delta_{IJ}, \quad \Delta_{IJ} = \begin{cases} 1, & disp = d \\ 0, & otherwise \end{cases}$$

$V_{jd}$  reprezintă valoarea din acumulatorul *V disparity* ce se incrementează la detecția unui pixel cu disparitatea  $d$  de pe rândul  $j$ .

Suprafața drumului se extrage folosind spațiul *V disparity* ca și o dreaptă parametrică. Astfel acest pas se rezumă la găsirea parametrilor dreptei din imagine ce conține cele mai multe puncte. Dreapta parametrică a drumului are ecuația  $y = mx + b$ . În Figura 12 avem prezentată o imagine în care a fost aplicat algoritmul menționat pentru a extrage suprafața drumului.



Figura 12. Imagine în care a fost extrasa suprafața drumului

Acumulatorul *U disparity* va fi utilizat pentru extragerea detaliilor despre coordonatele orizontale între care se află obstacolele. De asemenea, din spațiul *U disparity* se vor extrage și informații cu privire la intervalul disparităților în care se află obiectul.

Pentru calculul acumulatorului se va parcurge vectorul de puncte ce conține informații de disparitate și se va incrementa poziția corespunzătoare în matrice conform formulei de mai jos:

$$u_{di} = \sum_{j=0}^{rows} \Delta_{ij}, \Delta_{ij} = \begin{cases} 1, & disp = d \\ 0 & otherwise \end{cases}$$

În aceasta ecuație  $U_{di}$  reprezintă valoarea din acumulatorul  $U$  disparity ce se incrementează la detecția unui pixel cu disparitatea  $d$  de pe coloana  $i$ .

Următorul pas din cadrul algoritmului presupune detectarea coordonatelor orizontale ale punctelor de interes. Pentru a realiza acest lucru se va utiliza acumulatorul  $U$  disparity. Deoarece acest acumulator a fost creat prin incrementarea punctelor cu aceeași disparitate de pe aceleași linii în cadrul acestuia o să obținem aglomerări de puncte pentru obiecte cu disparități în range-uri apropiate.

Pentru extragerea obiectelor (zone continue de puncte) exista două abordări:

- Algoritmul Hough pentru a detecta liniile după care aplicăm un algoritm pentru a găsi liniile ce se intersectează (pentru a le considera ca aparținând aceluiași obiect)
- Utilizarea unui algoritm de etichetare

Pentru rezolvarea acestei probleme am ales să implementăm un algoritm de etichetare cu *kernel* de  $3 \times 11$ . După găsirea regiunilor de interes acestea vor fi filtrate în funcție de mai multe criterii: dimensiuni, poziție etc.

Rezultatul algoritmului poate fi remarcat în Figura 13.

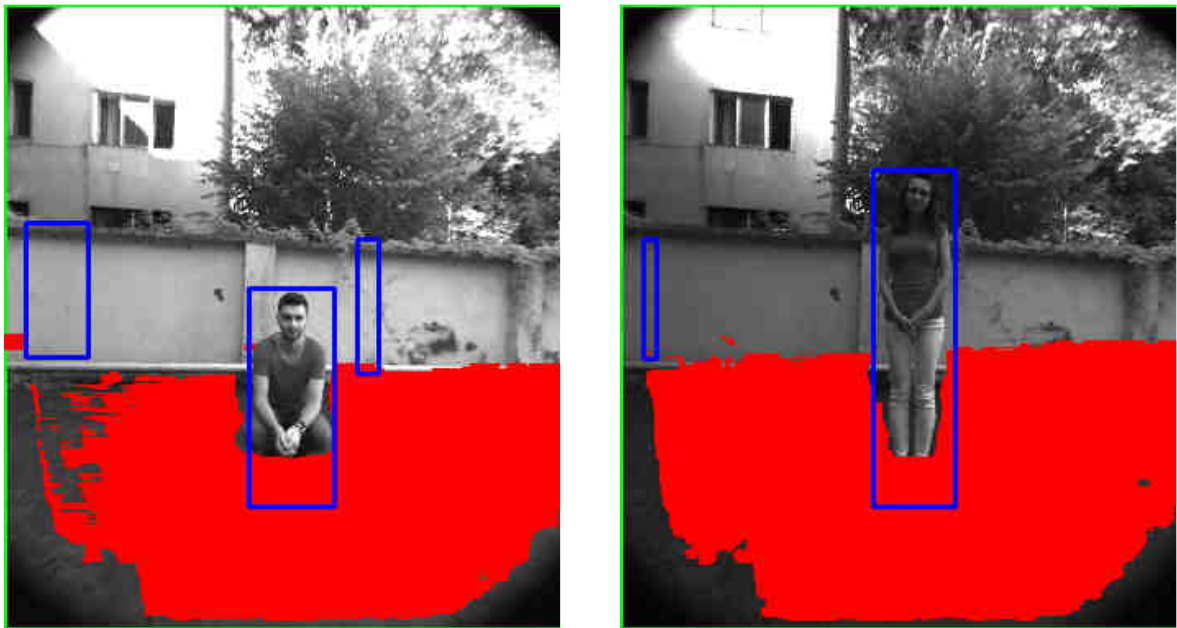


Figura 13. Identificarea obiectelor aflate deasupra solului, incluzând persoanele.

Poziția obiectelor împreună cu informația de profunzime sunt proiectate în camera a treia. Folosind poziția obiectelor reducem semnificativ spațiul și timpul de căutare al fețelor în imagini întrucât, fața se va căuta doar în regiunea de interes extrasă. De asemenea informația 3D ne oferă posibilitatea de a extrage trăsături bazate pe adâncime a căror combinație cu trăsăturile clasice extrase din imagini de intensitate oferă soluții mult mai robuste de detecție a fețelor și a trăsăturilor lor.



Pentru proiecția punctelor 3D în camera a treia s-au folosit parametrii intrinseci ai camerelor cât și parametrii extrinseci obținuți prin calibrare. Exemplu de proiecție în imagine a punctelor 3D este ilustrat în Figura 14.



Figura 14. Proiecția punctelor 3D în camera a 3-a

## 9. Proiectarea modelelor pentru geometria și poziția capului și pentru trăsăturile faciale.

În cadrul acestei etape am studiat modelele 3D existente pentru a descrie geometria facială și am ales un model adecvat care să permită urmărirea în timp real a poziției capului și măsurarea trăsăturilor faciale. De asemenea, am proiectat și implementat mai multe modele originale de urmărire a ochilor în timp real.

Am început prin a studia mai multe modele faciale 3D propuse în literatura de specialitate.

Modelul *Candide 3* [1] este un model 3D de modelare a fețelor umane, și care a fost dezvoltat la începutul anilor 90 pe baza metodologiei FACS (*Facial Action Coding Sistem*). Această metodologie permite măsurarea oricărei expresii faciale, prin exprimarea expresiei în acțiunile fundamentale ale unor mușchi sau grupe de mușchi (*Action Unit, AU*) și în intervalul temporal în care a apărut expresia. Intensitatea unui *Action Unit* (gradul de activare al mușchiului) poate fi indicată prin calificativele: A (*Trace – urmă*), B (*Slight – superficial*), C (*Pronounced – pronunțat*), D (*Extreme – extrem*) și E (*Maximum – maxim*).

*Candide 3* (Figura 15) este compus dintr-un număr redus de vârfuri (112), ceea ce permite generarea și reconstrucția rapidă a modelului cu resurse computaționale reduse. Fiecare vârf are o corespondență directă cu un *Facial Parameter (FP)* definit de standardul internațional Face and Body Animation (*FBA*) [2, 3]. Nu toate *FP*-urile sunt reprezentate în modelul *Candide*: de exemplu, nu sunt definite părul, urechile, limba și dinții. Modelul *Candide 3* definește 12 parametri prin care se poate deforma modelul neutru al feței către o față specifică, numiți *Shape Units (SU)*. Astfel, modelul generat poate reprezenta cel puțin fețele comune. Exemplu de unități de formă sunt: înălțimea capului, poziția sprâncenelor, poziția verticală a nasului etc. Modelul include mai multe *Animation Units*. Un *Animation Unit* este definit prin implementarea corespunzătoare a unuia sau mai multor *AU* în modelul

Candide. Cu alte cuvinte, *Shape Units* descriu forma particulară a feței subiectului, iar *Animation Units* sunt deformări ale formei neutre.

Mai recent, s-au dezvoltat modele faciale 3D mult mai complexe, generate cu ajutorul laserelor 3D. Modelul facial Basel Face Model (*BFM*) [4] a fost construit prin scanarea cu un laser 3D a fețelor a 100 de bărbați și 100 de femei. Modelul este definit de 53490 de vârfuri și de topologia asociată acestor vârfuri. În plus, fiecare vârf are asociate informații despre culoarea în acel punct printr-un vector tridimensional (r, g, b) – canalul roșu, albastru și verde.

Modelul Candide este frecvent utilizat în domeniul aplicațiilor de viziune artificială datorită simplității sale, care permite procesarea și recunoașterea expresiilor faciale în timp real. Datorită acestui avantaj, am hotărât să folosim modelul Candide pentru măsurarea și urmărirea trăsăturilor faciale.

Într-o primă etapă am dezvoltat o aplicație (Figura 16) care permite potrivirea manuală a modelului Candide 3 peste imagini faciale. Modelul 3D este încărcat dintr-un fișier *.wmf* și apoi este proiectat în imagine. Aplicația permite modificarea tuturor parametrilor care modelează forma (înălțimea capului, poziția verticală a ochilor, lățimea maxilarului etc.) și animațiile (ridicare de sprâncene, clipire, ridicare nas etc.) prin intermediul unor *slider-e*. De asemenea, se pot aplica transformările geometrice comune (translație, scalare și rotație) asupra modelului 3D. Principalul scop al acestei aplicații este înțelegerea în profunzime a modului în care ar trebui potrivit modelul 3D peste imagine: dacă există o anumită ordine în care ar trebui să fie efectuată potrivirea modelului și dacă anumite elemente faciale ar trebui să aibă prioritate în acest proces.

Din testele efectuate, am dedus că inițial ar trebui stabilite transformările geometrice ale modelului și parametrii de formă, iar ulterior ar trebui determinate animațiile care au fost aplicate modelului.

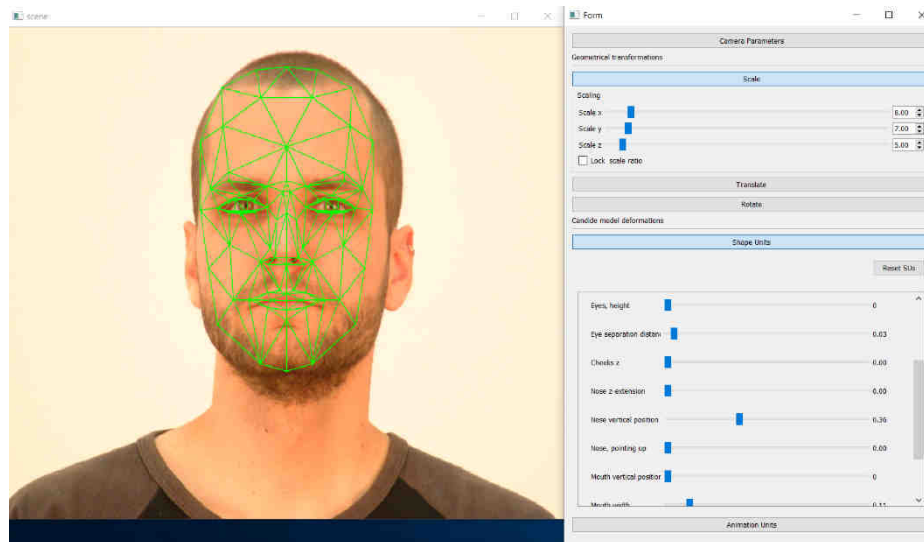


Figura 15. Potrivirea manuală a modelului Candide 3 peste o imagine

Întrucât unul din obiectivele principale ale proiectului este urmărirea ochilor în camere high speed, am proiectat și implementat mai multe modele simple, care să permită detectarea și urmărirea rapidă a ochilor.

Ochii sunt cele mai importante elemente ale feței, iar mișcările lor au un rol important în exprimarea stărilor emoționale și a proceselor cognitive.

Sistemele de urmărire a ochilor pot fi clasificate în *metode active* (care utilizează LED-uri cu lumină infraroșie, IR) și *metode pasive* (care se bazează pe lumina ambientală). Există

două tipuri generale de metode bazate pe lumină activă care pot fi exploatate folosind o anumită proprietate fizică a pupilei, care își modifică aparența în imaginea capturată în funcție de poziția iluminatorului IR: *pupila luminoasă (bright pupil)* și *pupila întunecată (dark pupil)*.

În lucrarea [7] este prezentată o trecere în revistă detaliată a tehnicilor de detecție și urmărire a ochilor, subliniindu-se atât problemele ridicate de acestea, cât și importanța lor în cadrul unui spectru larg de aplicații. Autorii propun o taxonomie formată din trei clase care se bazează pe modelul folosit pentru reprezentarea ochilor: *metode bazate pe forme*, *metode bazate pe aspect* și *metode hibride*.

*Metodele bazate pe forme* utilizează un model preliminar al geometriei ochiului și a texturii zonei din jurul său, precum și o măsură a similarității.

*Metodele bazate pe aspect* detectează și urmăresc ochii pe baza distribuției culorilor sau filtrează răspunsul regiunii ochiului. Aceste metode necesită volume mari de date de antrenament care reprezintă ochii în diferite condiții de iluminare și poziții ale feței.

În sfârșit, *metodele hibride* combină două sau mai multe abordări în vederea exploatării beneficiilor și a depășirii neajunsurilor acestora. Cu titlu de exemplu, framework-ul modelelor constrânse local (*Constrained Local models, CLM*) [8], o metodă de detecție a trăsăturilor faciale apărută recent și considerată foarte promițătoare, folosește un model bazat pe formă pentru a impune constrângeri asupra locului în care ar putea apărea fiecare trăsătură și un model complementar (*patch model*) pentru a descrie aparența trăsăturii. În [9] a fost propusă o metodă hibridă pentru extragerea trăsăturilor feței umane. Pentru aceasta sunt evaluați și combinați o serie de detectori ai trăsăturilor faciale individuale pe baza constrângerilor legate de formă. În final, rezultatele sunt rafinate folosind modele active de aparență (*Active Appearance Models, AAM*) reglate pentru cazurile speciale legate de muchii și de colțuri. În lucrarea [10] se prezintă o metodă în 3 etape de detecție a trăsăturilor faciale. În prima etapă, regiunea feței este localizată pe baza distanței Hausdorff dintre muchiile imaginii de intrare și un șablon de muchii faciale. În cea de-a doua etapă se folosește un model similar, dar mai mic, pentru ochi. În final, pozițiile pupilelor sunt rafinate folosind un perceptron multi-strat antrenat cu imagini centrate pe pupile. Alte metode localizează mai multe trăsături ale feței folosind descriptori de reper locali și un model predefinit al formei feței. În lucrarea [11] autorii propun o nouă abordare pentru localizarea trăsăturilor din imagini faciale. Metoda lor utilizează detectori locali pentru fiecare trăsătură și combină ieșirile lor cu un set de modele globale pentru regiunile parțiale calculate pe baza unui set etichetat de exemple. Pentru clasificarea fiecărei trăsături se utilizează *patch-uri* și un clasificator de tip *support vector machine (SVM)* combinat cu trăsături gray-scale de tipul *scale invariant feature transform (SIFT)*.

Prima metodă pe care am dezvoltat-o este o metodă rapidă și precisă de extragere a mai multor trăsături din jurul ochilor: centrul pupilei, raza irisului și forma externă a ochiului. Aceste trăsături sunt extrase folosind un algoritm cu mai multe etape. În prima etapă centrul pupilei este localizat folosind un detector rapid de simetrie circulară, iar raza irisului este calculată folosind proiecții radiale ale gradientului.

În cea de-a doua etapă forma externă a ochiului (a pleoapelor) este determinată cu ajutorul unui *framework* de eșantionare Monte Carlo bazat atât pe informațiile de culoare, cât și pe cele de formă. Designul soluției este prezentat în Figura 16.

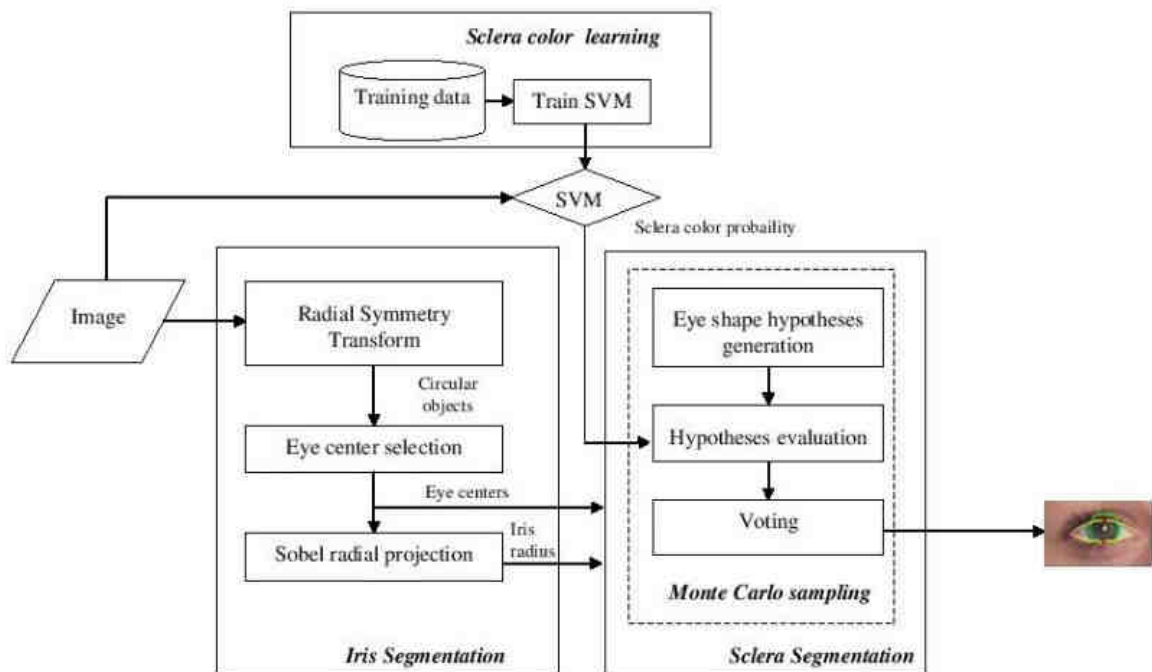


Figura 16. Designul soluției propuse pentru extragerea trăsăturilor ochilor

Prin pas al algoritmului este detecția fețelor din imagine și selectarea zonei de interes a ochilor pe baza unor constrângeri antropomorfe. Apoi, regiunile circulare din zona ochilor sunt detectate utilizând un detector de simetrii circulare: *Fast Radial Symmetry Transform* *FRST* [12]. *FRST* este o transformare a imaginii care utilizează gradientul imaginii pentru a determina rolul pe care fiecare pixel  $p$  îl are la simetria pixelilor vecini aflați la o distanță  $r$  de acest pixel. Această contribuție se calculează prin însumarea contribuțiilor magnitudinii și orientării pe direcția gradientului.

Candidații pentru centrelor irisurilor sunt determinați ca minime locale din transformata *FRST*, iar centrele irisurilor sunt selectate din candidații determinați la pasul anterior pe baza unor constrângeri geometrice.

Raza irisurilor se calculează utilizând derivata Sobel a imaginii pe o regiune în jurul ochilor pentru a accentua tranziția puternică dintre zona irisului și a sclerei. Pentru a elimina zgomotele din această regiune, pe imagine se aplică un filtru Gaussian și apoi primele  $k\%$  cele mai mici valori din derivata Sobel sunt ignorate. Pentru fiecare rază candidat din intervalul  $[r_{min}, r_{max}]$ , se calculează o proiecție radială prin însumarea valorilor gradientului care sunt la o distanță  $r$  de centrul irisului și cu o deschidere angulară între  $[\theta_{min}, \theta_{max}]$ . Această proiecție a imaginii atinge valoarea maximă la granița dintre regiunea sclerei și a irisului.

Forma ochilor este reprezentată de două parabole, una pentru pleoapa de sus și una pentru pleoapa de jos (Figura 18 (a)). Ca să potrivim o formă ipotetică la o imagine, luăm în considerare culoarea sclerei ochilor. Regiunea din jurul ochiului este transformată într-un spațiu probabilistic utilizând algoritmi de învățare automată pentru a determina probabilitatea unui pixel de a fi un pixel din regiunea sclerei. Am antrenat un clasificator de tipul Support Vector Machine (*SVM*) pe regiuni (*patch*) din zona ochilor pe imagini dintr-o bază de date publică. Trăsăturile utilizate pentru învățare sunt: canalul de *Hue* din spațiul de culoare *HSV* și canalele *O1* și *O2* din spațiul de culoare *RGB* opus (*RGB opponent*).

Pentru a evalua scorul unei ipoteze am analizat două mulțimi de pixeli la o distanță  $\Delta$  de conturul ipotezei curente: pixelii pozitivi ( $p_+$ ), care reprezintă pixelii din regiunea sclerei și pixelii negativi ( $p_-$ ) care reprezintă pixelii din zona pielii și a genelor. Procesul de potrivire

(doar pentru pleoapa de sus) este ilustrat în figura 17 (b). Scorul de potrivire al unei ipoteze este definit ca:

$$S = \frac{\alpha \sum \Delta P_+ - \beta \sum \Delta P_-}{\alpha + \beta}$$

unde  $\alpha$  și  $\beta$  ( $\alpha + \beta = 1$ ) sunt două ponderi care determină influența pixelilor pozitivi și respective negative.

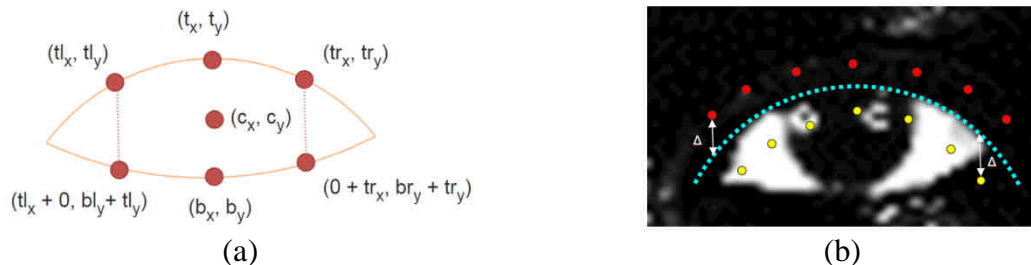


Figura 17. (a) Modelul utilizat pentru a modela forma ochiului. (b). Procesul de potrivire al unei ipoteze: linia punctată reprezintă conturul pleoapei de sus al unei ipoteze, iar pixelii pozitivi sunt marcați cu galben și pixelii negativi cu roșu.

Soluția a fost testată pe 3 baze de date disponibile publicului: BIO-ID [10], LFWD[11] și Park [13]. Rezultatele pe diferite imagini din baza de date Park sunt prezentate în Figura 19.

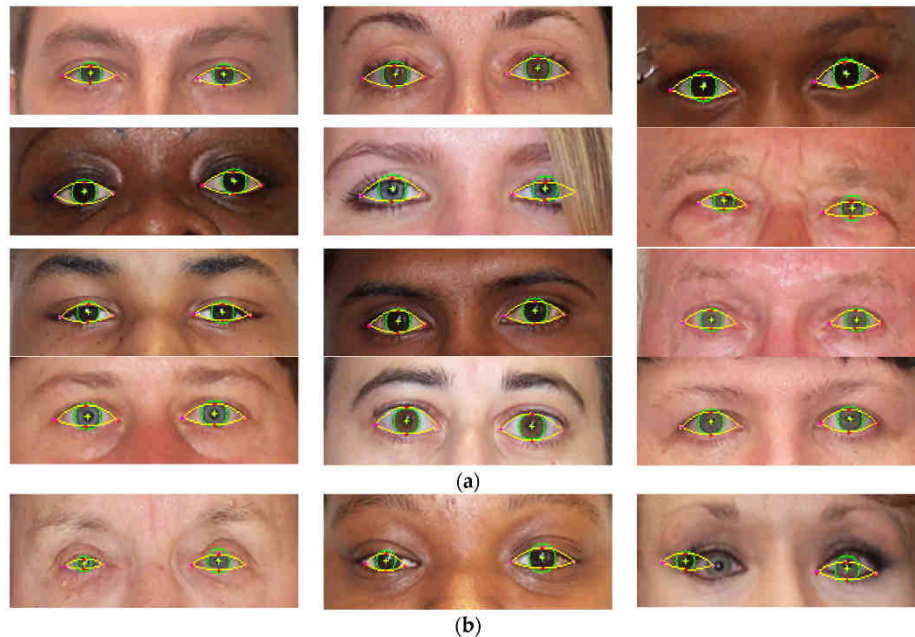


Figura 18. Rezultatele metodei de segmentare a ochilor în imagini color din baza de date Park. (a) Rezultate corecte. (b) Rezultate eronate

Rezultatele experimentale efectuate arată că în 96.30% din imagini, aproximarea cea mai slabă (dintre ochiul stâng și ochiul drept) pentru centrul pupilei este mai mică decât raza pupilei. În medie, diferența medie dintre raza detectată a irisului și raza reală normalizată cu raza reală este 0.0991; cu alte cuvinte, eroarea medie a irisului este mai mică decât 10% din raza irisului. Pe imaginile testate, magnitudinea corespunzătoare acestui procent este de 1-2 pixeli.

Principalul dezavantaj al metodei prezentate anterior este faptul că se bazează pe informații de culoare și nu este robustă la variațiile de iluminare. Astfel, următorul pas a fost modificarea soluției astfel încât aceasta să fie invariabilă la variațiile de iluminare.

Metoda prezentată anterior a fost publicată într-un jurnal de specialitate [19].

Pentru a elimina aceste dezavantaje ale algoritmului, am propus o nouă metodă bazată pe filtre de particule și care utilizează mai multe informații, nu doar informații despre culoare, pentru măsurarea gradului în care o particulă se potrivește în imaginea de intrare.

Filtrele de particule [5] reprezintă tehnici de implementare a filtrelor Bayes recursive prin eșantionare Monte Carlo; aceste metode au fost deja aplicate cu succes în probleme de localizare, urmărirea obiectelor și segmentarea imaginilor medicale. Filtrul urmărește  $N$  valori discrete, numite particule sau mostre ( $X_i$ ); fiecare particulă are o pondere ( $\pi_i$ ) care îi este asociată. Suma tuturor ponderilor este egală cu 1. Estimările filtrului de particule sunt calculate pe baza acestor mostre și ponderi.

Primul pas al acestui algoritm îl constituie *reeșantionarea (resampling)*. Mulțimea particulelor este transformată într-o nouă mulțime de particule neponderate prin efectuarea a  $N$  extrageri aleatoare cu înlocuire conform ponderilor (principiul „supraviețuirii celui mai înzestrat”, *survival of the fittest*). O particulă cu o valoare mai mare a ponderii asociate ar putea fi selectată de mai multe ori în noua populație, în timp ce o particulă care are o valoare mai mică a ponderii asociate s-ar putea să nu fie aleasă deloc. Rezultă o nouă mulțime de particule de concentrație inegală care aproximează aceeași funcție de densitate ca și mostrele precedente.

Următorul pas al aplicării filtrului de particule îl constituie *predicția*, în care potrivirea (*likelihood*) fiecărei stări curente posibile este calculată în ipoteza că fiecare particulă neponderată descrie starea trecută a sistemului. Toate particulele selectate în pasul de reeșantionare sunt supuse unei transformări deterministe – dictate de tranziția sistemului – și unei difuziuni stohastice care ia în considerare evenimentele aleatoare care pot influența sau modifica starea sistemului.

*Măsurarea / actualizarea* reprezintă pasul final al algoritmului. Acum se pune problema de a obține funcția de densitate de probabilitate posterioară pe baza celei anterioare și a măsurătorilor. Densitatea de probabilitate anterioară este exprimată prin distribuția particulelor neponderate, iar funcția de densitate de probabilitate posterioară este obținută prin simpla atribuire de ponderi particulelor în funcție de potrivirea (*likelihood*) observațiilor.

La începutul algoritmului, mulțimea de particule este construită prin generarea aleatoare a particulelor pe baza distribuției uniforme de probabilități, astfel încât starea inițială acoperă întregul spațiu de căutare. În etapele următoare, până la îndeplinirea unei anumite condiții de terminare, se repetă în mod iterativ pașii canonici ai filtrului de particule: *reeșantionarea, predicția și măsurarea / actualizarea*.

O particulă care descrie forma ochiului este complet descrisă de următorii parametri (Figura 20).

$$\left[ \begin{array}{ll} (x, y) & \text{centrul ochiului} \\ w & \text{lățimea ochiului} \\ h & \text{înălțimea ochiului} \\ \alpha & \text{orientarea ochiului} \\ d_{corners} & \text{poziția relativă pe axa } Oy \text{ a colțurilor ochiului} \end{array} \right]$$

, unde  $(x, y)$  sunt coordonatele centrului ochiului,  $w$  și  $h$  sunt lățimea și înălțimea ochiului,  $\alpha$  este unghiul dintre colțul intern și extern al ochiului, iar  $d_{corners}$  determină poziția verticală a colțurilor ochiului față de înălțimea ochiului. Valoarea  $d_{corners}$  este definită ca procent din înălțimea ochiului.

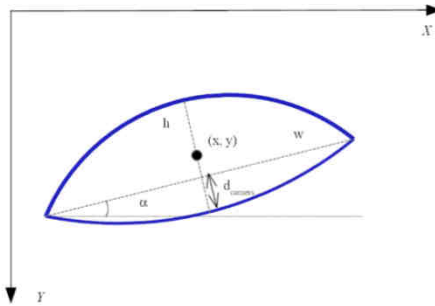


Figura 19. Reprezentarea vizuală a particulei ochiului. Forma ochiului este reprezentată cu albastru. Punctul  $(x, y)$  este centrul de greutate al formei ochiului,  $\alpha$  este orientarea ochiului,  $y_{corners}$  este coordonata  $y$  a colțurilor ochiului.

Pentru a determina în ce măsură o particulă se potrivește cu imaginea de intrare, am utilizat mai multe metrice pentru a discrimina între mulțimea de particule:

- orientarea gradientului: dacă un punct de muchie dintr-o imagine face parte dintr-un contur, atunci gradientul într-un punct de muchie al acestui contur este perpendicular pe tangenta la acest contur în punctul respectiv. Pentru a măsura această metrică de potrivire am evaluat valoarea medie a cosinusului diferenței dintre unghiul gradientului și cel al tangentei pe conturul ipotetic al formei ochiului;

- canalul de saturație: porțiunea sclerei este în mod normal cea mai deschisă din zona feței și are o culoare foarte apropiată de alb. Astfel, regiunea corespunzătoare sclerei ar trebui să aibă cele mai mici valori ale saturației;

- transformata de simetrie *FRST* [12]: irisul este cuprins în interiorul ochiului, astfel că în forma ipotetică a ochilor ar trebui să avem o regiune simetrică circulară. Pentru a evalua această metrică am utilizat transformata de simetrie radială;

- valorile colțurilor: întrucât urmărim să găsim colțurile ochilor, este natural să analizăm și răspunsul unui detector consacrat de colțuri în imagini.

Pe baza acestor trăsături am definit mai multe distanțe care sunt fuzionate într-un produs de funcții Gaussiene pentru a calcula ponderea particulelor.

Performanța metodei propuse a fost evaluată pe imagini din baza de date UbiPr [14], care conține imagini perioculare captate în condiții și scenarii de achiziție necontrolate. Fiecare imagine din setul de date a fost adnotată manual de către autori cu centrul irisului, colțul intern și colțul extern al ochiului și mai mulți markeri pe sprâncene. Sunt de asemenea disponibile adnotări referitoare la condițiile de captură (unghiul direcției de privire și unghiul de poză, prezența ochelarilor, ocluzii date de păr etc.).

În Figura 20 sunt prezentate niște exemple de detecție ale ochilor pe imagini din baza de date UbiPr.

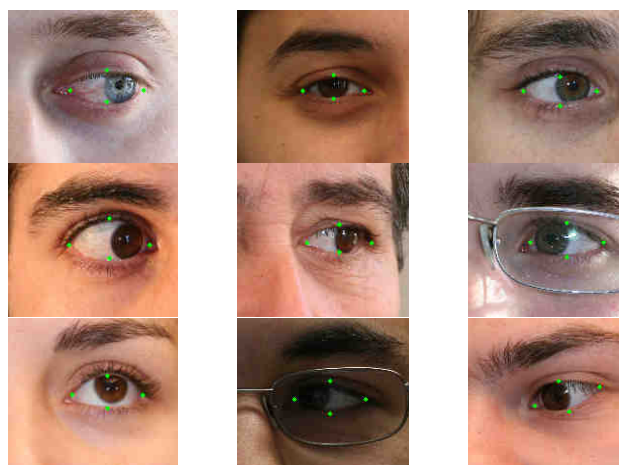


Figura 20. Rezultatele metodei de detecție ale colțurilor ochilor în imagini din baza de date UbiPr

Am evaluat *in extenso* soluția propusă pe imaginile din *dataset*.

Eroarea medie pentru detecția colțurilor ochiului este de 21.359 pixeli (atât pentru regiunea nazală, cât și pentru regiunea temporală a ochiului), cea ce reprezintă aproximativ 9.5% din lățimea ochiului. Folosind constrângeri antropomorfe putem conchide că eroarea medie este de mai puțin 50% din raza irisului. Această metodă a fost prezentată în cadrul unei conferințe de specialitate [20].

Următorul pas a fost dezvoltarea unei metode care să permită urmărirea ochilor în cadrele video. Soluția propusă folosește o abordare de la grosier la fin (*coarse-to-fine*) pentru a detecta și urmări multiple trăsături ale ochilor: centrul irisului, conturul ochiului și clipirile.

Pașii soluției sunt prezentați în Figura 21. Soluția utilizează trei filtre de particule în paralel pentru a urmări ochii: primul filtru de particule este folosit pentru a determina pozițiile aproximative ale irisurilor. Alte două filtre de particule sunt utilizate pentru a determina și a urmări conturul fiecărui ochi pe baza estimării obținute de la primul filtru de particule.

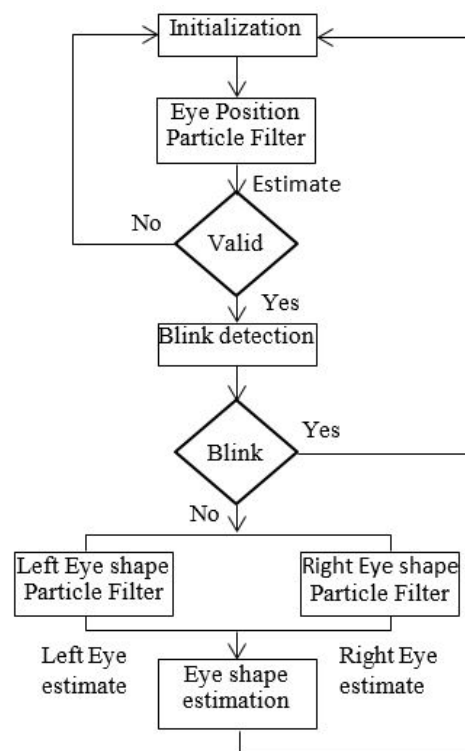


Figura 21. Urmărirea ochilor în cadre video

În prima etapă se utilizează un filtru de particule pentru a estima la modul grosier poziția centrilor irisurilor și orientarea ochilor (unghiul dintre linia care unește cei doi ochi și axa orizontală). În primul cadru video, filtrul de particule este inițializat: fața este detectată în imaginea de intrare și spațiul de căutare este în mod uniform acoperit prin generarea aleatoare de particule în întregul spațiu de intrare.

În cadrele video următoare, se aplică iterativ filtrul de particule pentru a actualiza estimarea poziției ochilor.

În mod periodic, din  $t$  în  $t$  cadre, estimarea obținută este supusă unui test de validitate: fața este detectată în cadrul curent și se verifică dacă pozițiile estimate ale ochilor se află în zona aproximativă a ochilor (regiunea din jumătatea de sus a feței). Dacă estimarea nu mai este validă, atunci urmărirea este re-inițializată.

Pentru potrivirea unei particule am analizat simetria circulară a irisului: s-a aplicat pe fiecare cadru video o transformare *Fast Radial Symmetry Transform* [12] care calculează pe



baza derivatelor de gradul 1 ale imaginii simetria circulară observabilă în fiecare pixel. Ponderea unei particule este dată de media răspunsului transformării *FRST* în pixelii corespunzători centrelor irișilor.

Dacă estimarea este validă, ea este analizată pentru a determina dacă ea corespunde stării închise sau deschise a ochiului. S-a utilizat o metodă simplă pentru detecția clipirilor: se analizează schimbarea magnitudinii răspunsului circular de la cadrul anterior la cadrul curent. Dacă această magnitudine a scăzut cu mai mult de un anumit prag, considerăm că ochiul este închis. În cazul în care am detectat că ochiul este deschis, celelalte două filtre de particule utilizate primesc informații despre noua poziție estimată a irisului și sunt utilizate pentru a determina forma exactă a ochilor.

Aceste două filtre de particule utilizează același vector de stare pentru a descrie o particulă ca cel ilustrat în Figura 19. Pentru potrivirea particulei la imaginea de intrare s-a utilizat Transformata Distanță (*Distance Transform*) [15]. Transformata distanță este o reprezentare a imaginii în care fiecare pixel indică distanța față de cel mai apropiat punct de muchie. Având o particulă care reprezintă forma unui ochi, scorul ei de potrivire se calculează prin suprapunerea formei ipotetice a ochiului peste Transformata distanță.

Metoda a fost intens evaluată pe baza de date Cohn-Kanade [16, 17]. Erorile medii obținute sunt afișate în Tabelul 8.1.  $D$  este distanța Euclidiană în pixeli dintre locația detectată și locația reală (*ground truth*) și  $\tilde{D}$  este distanța euclidiană normalizată.

Tabelul 8.1. Erorile medii pe baza de date Cohn-Kanade

	Colțul ochiului din regiunea temporală	Colțul ochiului din regiunea nazală	Limita superioară a pleoapei	Limita inferioară a pleoapei	Centrul irisului
$D$ (pixeli)	4.88	6.42	3.15	2.69	5.3
$\tilde{D}$ (%)	0.11	0.14	0.07	0.06	0.05

Rezultatele obținute pe niște imagini din baza de date Cohn-Kanade sunt ilustrate în Figura 22. Această metodă a fost trimisă spre publicare la o conferință de specialitate.

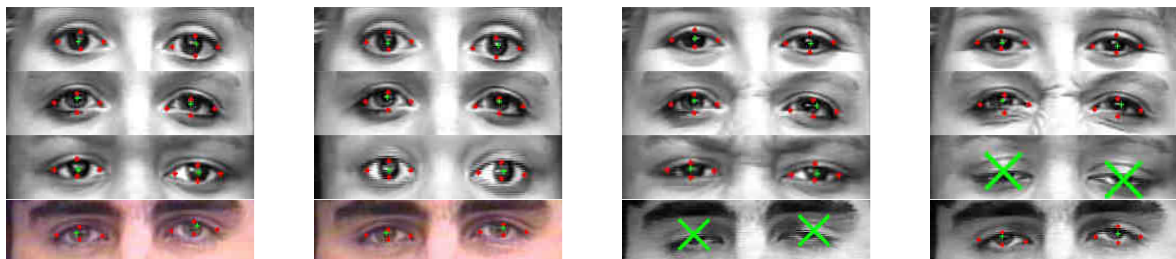


Figura 22. Rezultatele algoritmului pe imagini din baza de date Cohn Kanade

Pentru a rezuma, în cadrul acestei activități am studiat modelele 3D existente pentru a modela fețele umane. Ne-am focalizat asupra a două modele disponibile public: modelul Candide 3 și Basel Face Model. Fiindcă obiectivul care trebuie îndeplinit este urmărirea capului în timp real, am ales să utilizăm modelul facial Candide 3. Ulterior am dezvoltat o aplicație prin care am suprapus manual modelul facial la o imagine; această aplicație ne-a permis să studiem modul în care un om ar potrivi modelul facial pe o imagine și să determinăm elementele principale care influențează acest proces.

Apoi am proiectat și implementat mai multe modele de detecție și urmărire a ochilor în imagini. Metodele dezvoltate au aplicabilitate într-o varietate de domenii: interacțiune om calculator, sisteme de monitorizare ale șoferului, realitate augmentată, biometrie etc. Prin detectarea precisă a colțurilor ochilor și a pleoapelor se poate determina regiunea sclerei.

Recent, s-a dovedit că segmentarea sclerei are o importanță deosebită în contextul măsurătorilor biometrice neconstrânse, în lumină vizibilă, ale irisului și sclerei. Există mai multe lucrări care au început să trateze această problemă. În plus, s-au organizat competiții de segmentarea sclerei pe bază de *benchmarking* [18] în vederea evaluării progreselor recente din acest domeniu și pentru a atrage atenția cercetătorilor asupra lui.

## 10. Proiectarea și implementarea algoritmilor de urmărire a geometriei și poziției capului faciale.

În această etapă a fost dezvoltată o aplicație care potrivește automat modelul Candide 3 pe imagini faciale utilizând filtre de particule. Filtrele de particule au fost descrise în detaliu în Secțiunea 8.

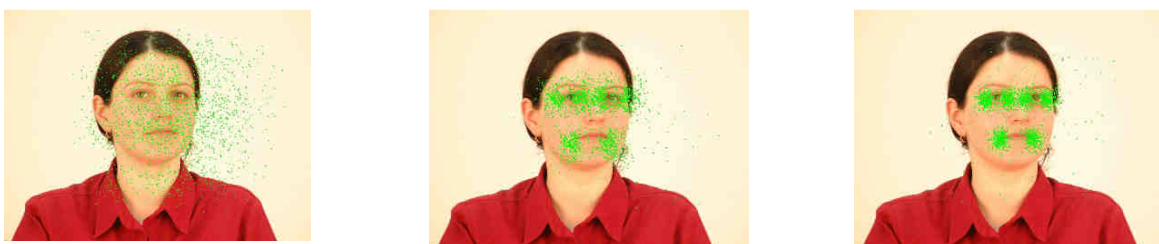
O particulă este descrisă de următorii parametrii:

$$\left[ \begin{array}{l} x, y, z - \text{translatia} \\ s - \text{factorul de scalare} \\ r_x, r_y, r_z - \text{rotația} \\ \text{shape} - \text{parametrii de formă} \end{array} \right]$$

,unde  $(x, y, z)$  este un vector tridimensional care reprezintă translația modelului,  $s$  este factorul de scalare,  $(r_x, r_y, r_z)$  este vectorul de rotație și  $\text{shape}$  este un tablou unidimensional care conține ponderile unităților de formă (*Shape units*) din modelul Candide. Pe baza acestor informații se generează modelul Candide 3 și apoi fiecare punct al modelului este proiectat în imagine. Pentru a estima distanța (pe axa Z) a modelului, am ținut cont de distanța inter-pupilară medie a omului și de distanța focală a camerei cu care au fost capturate fotografiile.

Pentru potrivirea unei particule peste imaginea de intrare am evaluat doar două metrici legate de poziția ochilor și de colțurile gurii. Pentru a evalua potrivirea în zona ochilor am evaluat distanța euclidiană dintre colțurile ochilor modelului unei particule și poziția ochilor în imagine. În prealabil, pe imagine am aplicat metoda de determinare a colțurilor ochilor pe care am dezvoltat-o (Secțiunea 8). Potrivirea colțurilor gurii se face prin analiza răspunsul unui detector de colțuri în zona gurii.

Filtrul de particule rulează iterativ până când valoarea estimată de la o iterație la alta a algoritmului este mai mică decât un prag  $\epsilon$ . În Figura 23 este ilustrată evoluția filtrului de particule și estimarea obținută.



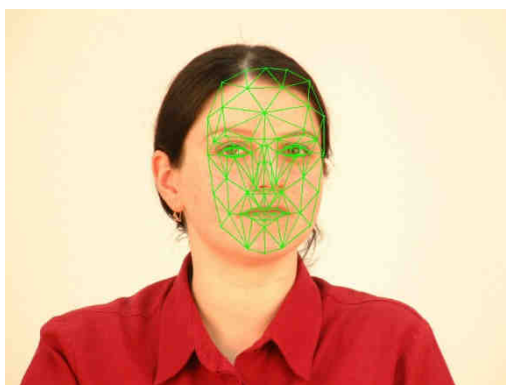


Figura 24. Evoluția populației filtrului de particule și estimarea obținută

Estimarea modelul feței se face prin calculul mediei ponderate a celor mai bune 30% particule; fiecare particulă aleasă este ponderată cu *weight*-ul ei. Din această estimare se poate determina poziția și orientarea capului ( prin vectorii  $(x, y, z)$  și  $(r_x, r_y, r_z)$  ), informații despre forma și dimensiunea feței, precum și poziția în imagine a trăsăturilor faciale.

Soluția a fost testată pe imagini dintr-o bază de date cu imagini faciale cu diferite orientări ale capului [6].

În viitor, ne propunem să extindem această metodă prin adăugarea parametrilor de animație ai modelului Candide-3. În versiunea curentă, se presupune că fața are o expresie neutră. Metoda poate fi cu ușurință extinsă astfel încât să detecteze poziția capului și să urmărească trăsăturile faciale în cadre video.

## 11. Diseminarea rezultatelor preliminare

Pentru a face publice cele mai importante realizări tehnice pe care le-am obținut am elaborat și am trimis spre publicare mai multe articole în reviste și conferințe de specialitate.

Articole publicate:

- Borza, D., Darabant, A. S., & Danescu, R. (2016). Real-Time Detection and Measurement of Eye Features from Color Images. *Sensors*, 16(7), 1105. [ISI]
- M. P. Muresan, S. Nedevschi & R. Danescu. (2016) "Patch warping and local constraints for improved block matching stereo correspondence," *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, Romania, 2016, pp. 321-327, doi: 10.1109/ICCP.2016.7737167

Articole trimise spre publicare și acceptate (*accepted*):

- Borza, D. & Danescu, R. (2016). Eye Shape and Corners Detection in Periocular Images Using Particle Filters. In *Proceedings of the 12th International Conference on Signal Image Technology & Internet Based Systems (SITIS)*, 28 Nov – 1 Dec, Naples, Italy.

Articole trimise spre publicare, in evaluare (*submitted*):

- Borza, D., Darabant, A. S., & Danescu, R. "Fast Eye Tracking and Feature Measurement Using a Multi-Stage Particle Filter". In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 27 Febr – 1 March 2017.
- M. P. Muresan, S. Nedevschi & R. Danescu. "A Multi Patch Warping Approach For Improved Stereo Block Matching". In *Proceedings of the 12th International Joint Conference*

on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP),  
27 Febr – 1 March 2017.

## Bibliografie

- [1] Jörgen Ahlberg, "CANDIDE-3 - An Updated Parameterised Face," 2001.
- [2] VisageTechnologies. <http://www.visagetechnologies.com>, [Online], (2016, Nov.)  
<http://www.visagetechnologies.com/uploads/2012/08/MPEG-4FBAOverview.pdf>
- [3] Igor Pandzi and Robert Forchheimer, MPEG-4 Facial Animation: The Standard, Implementation And Applications.: John Wiley & Sons, Inc., 2003.
- [4] Volker Blanz and Thomas Vetter, "A Morphable Model for the synthesis of 3d faces," in *26th annual conference on Computer graphics and interactive techniques*, New York, 1999.
- [5] Michael Isard and Andrew Blake. "Condensation - conditional density propagation for visual tracking". *International Journal of Computer Vision*, vol. 29, pp. 5–28, 1998, DOI: 10.1023/A:1008078328650.
- [6] A. Kasiński, A. Florek and A. Schmidt "THE PUT FACE DATABASE" - *Image Processing & Communications Volume 13, Number 3-4*, 59-64, 2008.
- [7] D.W. Hansen and Ji, Q. "In the eye of the beholder: A survey of models for eyes and gaze". *IEEE Trans. Pattern Anal. Mach. Intell.* 2010, 32, 478–500.
- [8] D. Cristinacce and T. Cootes. "Automatic feature localisation with constrained local models". *Pattern Recognit.* 2008, 41, 3054–3067.
- [9] D. Cristinacce, T. Cootes and I.M Scott. "A Multi-Stage Approach to Facial Feature Detection". In *Proceedings of the British Machine Vision Conference*, Kingston, UK, 7–9 September 2004.
- [10] O. Jesorsky, K.J. Kirchberg and R.W. Frischholz. "Audio- and Video-Based Biometric Person Authentication". In *Proceedings of the Third International Conference, AVBPA 2001*, Halmstad, Sweden, 6–8 June 2001; Springer: Berlin, Germany; Heidelberg, Germany, 2001; pp. 90–95.
- [11] P. Belhumeur, D. Jacobs, D. Kriegman and N. Kumar. "Localizing Parts of Faces Using a Consensus of Exemplars". In *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 20–25 June 2011.
- [12] G. Loy and E. Zelinsky. "A Fast Radial Symmetry Transform for Detecting Points of Interest". In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, 28–31 May 2002; p. 358.
- [13] M. Minear and D.C. Park. "A lifespan database of adult facial stimuli". *Behav. Res. Methods Instrum. Comput.* 2004, 36, 630–633.
- [14] Hugo Proença, Sílvio Filipe, Ricardo Santos, João Oliveira and Luís A. Alexandre. "The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-The-Move and At-A-Distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August, 2010, vol. 32 (8), pp. 1529-1535, DOI: <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.66>.
- [15] R. Fabbri, L. Da, J.C. Torelli and O.M. Bruno. "2D euclidean distance transform algorithms: A comparative survey". *ACM Comput. Surv.* 2008, 40, 1–44.
- [16] T. Kanade, J. F Cohn and Y. Tian. "Comprehensive database for facial expression analysis". *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 46-53.
- [17] P. Lucey, J. F Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews. "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression". *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, San Francisco, USA, 94-101.
- [18] Sclera Segmentation Benchmarking Competition 2015. [Online] (Nov. 2016),  
<http://www.ict.griffith.edu.au/conferences/btas2015/index.html>.
- [19] D. Borza, A. S. Darabant and Radu Danescu. "Real-Time Detection and Measurement of Eye Features from Color Images". *Sensors*, 16 (7), 1105.
- [20] D. Borza and R. Danescu. "Eye Shape and Corners Detection in Periocular Images Using Particle Filters". In *Proceedings of the 12th International Conference on Signal Image Technology & Internet Based Systems (SITIS)*, 28 Nov – 1 Dec, Naples, Italy.