Optimizarea performanțelor sistemelor biometrice prin selecția punctelor de operare în analiza ROC

Sorin SOVIANY[1], Sorin PUŞCOCI[1]

Rezumat. *Articolul prezintă un model practic pentru evaluarea şi optimizarea performanțelor sistemelor biometrice, model care are în vedere adaptarea nivelelor de securitate la constrângerile de performanță şi de costuri impuse de aplicațiile utilizatorilor. Analiza ROC (Caracteristica de operare a receptorului) aplicată în cazul sistemelor biometrice conduce la stabilirea unui număr de puncte de operare, în funcție de pragurile de securitate/utilizabilitate impuse de aplicația client. Selecția punctului optim de operare permite adaptarea performanței sistemului biometric la constrângerile specifice ale aplicației.*

Cuvinte cheie: *analiză ROC, puncte de operare, optimizare.*

Abstract. *The paper presents a practical framework for the biometric systems performances assessment and optimization; this approach is looking to adjust the provided security levels to the users applications-specific performance and costs-related constraints. The ROC (Receiver Operation Characteristic) Analysis for biometric systems provides some operating points according to the security/useness thresholding given by the client application. The suitable operating point selection allows to adjust the biometric system performance to the application specific constraints.*

Keywords: *ROC analysis, operating points, optimization.*

1. INTRODUCERE

Sistemele actuale de securitate bazate pe tehnologii biometrice prezintă performanțe care depind de o multitudine de factori obiectivi şi subiectivi, legați atât de calitatea şi precizia algoritmilor proiectați, dezvoltați şi implementați, cât şi de condițiile externe pentru achiziția şi înregistrarea datelor biometrice [1, 2] În plus, performanțele obținute în aplicațiile biometrice de verificare dar mai ales în cele de identificare a persoanelor, indiferent de tehnologiile biometrice utilizate, variază în funcție şi de cerințele aplicațiilor, respectiv de pragurile de securitate/ utilizabilitate admise sau fixate. Această variabilitate este specifică sistemelor de securitate bazate pe tehnologii biometrice, în condițiile unei variabilități inerente a şabloanelor biometrice generate la fiecare tentativă de autentificare, şi este una dintre diferențele majore față de alte clase de sisteme de securitate.

Una dintre abordările tipice în evaluarea şi optimizarea performanțelor pentru aplicații de recunoaştere de paternuri şi probleme de clasificare (inclusiv de date biometrice) constă în aplicarea unei strategii de analiză bazate pe reprezentarea curbelor ROC (Receiver of Operation Characteristic), folosind indicatori de performanță cum ar fi rata rezultatelor fals pozitive sau corect negative, de exemplu, în funcție de anumite valori prag specifice aplicațiilor.

[1] Institutul Național de Studii şi Cercetări pentru Comunicații – I.N.S.C.C, Bucureşti.

Utilizarea unei strategii de analiză ROC în cazul sistemelor biometrice cu selecția punctului sau punctelor optime de operare (pentru praguri de securitate/ utilizabilitate setate la nivel de aplicație) permite adaptarea rapidă a performanței unui sistem biometric la constrângerile specifice ale aplicației.

În continuare articolul este organizat astfel. Secțiunea II prezintă fundamente teoretice suport pentru aplicarea analizei ROC în cazul sistemelor biometrice. Secțiunea III este o analiză de caz pentru optimizarea performanței unui sistem biometric prin selecția punctului sau punctelor de operare în acord cu cerințele aplicației. Secțiunea IV prezintă concluzii rezultate din aplicarea metodei de analiză şi optimizare bazate pe analiză ROC pentru caracterizarea şi îmbunătățirea performanțelor sistemelor biometrice.

2. FUNDAMENTE TEORETICE PRIVIND METODA DE OPTIMIZARE ŞI EVALUARE A SISTEMELOR BIOMETRICE FOLOSIND ANALIZA ROC

Pentru definirea metodei de evaluare şi optimizare a performanțelor sistemelor biometrice se are în vedere specificarea următoarelor elemente suport:

- indicatori de performanță utilizabili în analiza ROC;
- analiza ROC şi principii de aplicare în cazul sistemelor biometrice

2.1. Indicatori de performanță pentru sisteme biometrice ca sisteme de recunoaştere de paternuri, utilizabili în analiza ROC

Sistemele biometrice sunt, indiferent de tehnologiile utilizate (dispozitive de achiziție de date, algoritmi de pre-procesare şi procesare a datelor) **sisteme de recunoaştere a paternurilor**.

Componentele funcționale principale sunt, în conformitate cu arhitectura generică din figura 1, următoarele [3]:

- *blocul de achiziție de date:* unul sau mai mulți senzori pentru captura datelor;
- *blocul de pre-procesare:* realizează transformări primare ale datelor achiziționate;
- *blocul de extragere a caracteristicilor:* generează caracteristicile utile;
- *blocul de selecție a caracteristicilor:* elimină informațiile redundante şi mai puțin relevante;
- *blocul de selecție şi antrenare a modelului de clasificare:* alegerea modelului și antrenarea clasificatorului;
- *blocul de evaluare* care stabilește performanța de generalizare a sistemului proiectat**.**

Dacă performanța nu îndeplinește cerințele aplicației, modelul este optimizat suplimentar pentru adaptarea la nivelul de precizie dorit. Rezultatul sau decizia se transferă către aplicație [4, 5].

Funcția de bază care fundamentează decizia privind verificarea sau stabilirea identității unei persoane în autentificarea biometrică este cea de clasificare. Indiferent de **natura modelului de clasificare** aplicat ca parte a sistemului biometric proiectat:

- ***clasificare bazată de distanță***, în care se calculează *indicatori de similaritate* prin comparație directă între vectorii de caracteristici generați în stadiul de pre-procesare şi şabloanele biometrice de referință;
- ***clasificare supervizată*** folosind *modele antrenate cu seturi de date disponibile*, şi în care clasificarea se bazează pe *capacitatea modelului de a „învăța" din datele biometrice de referință* existente, nu pe comparația directă a vectorilor de caracteristici.

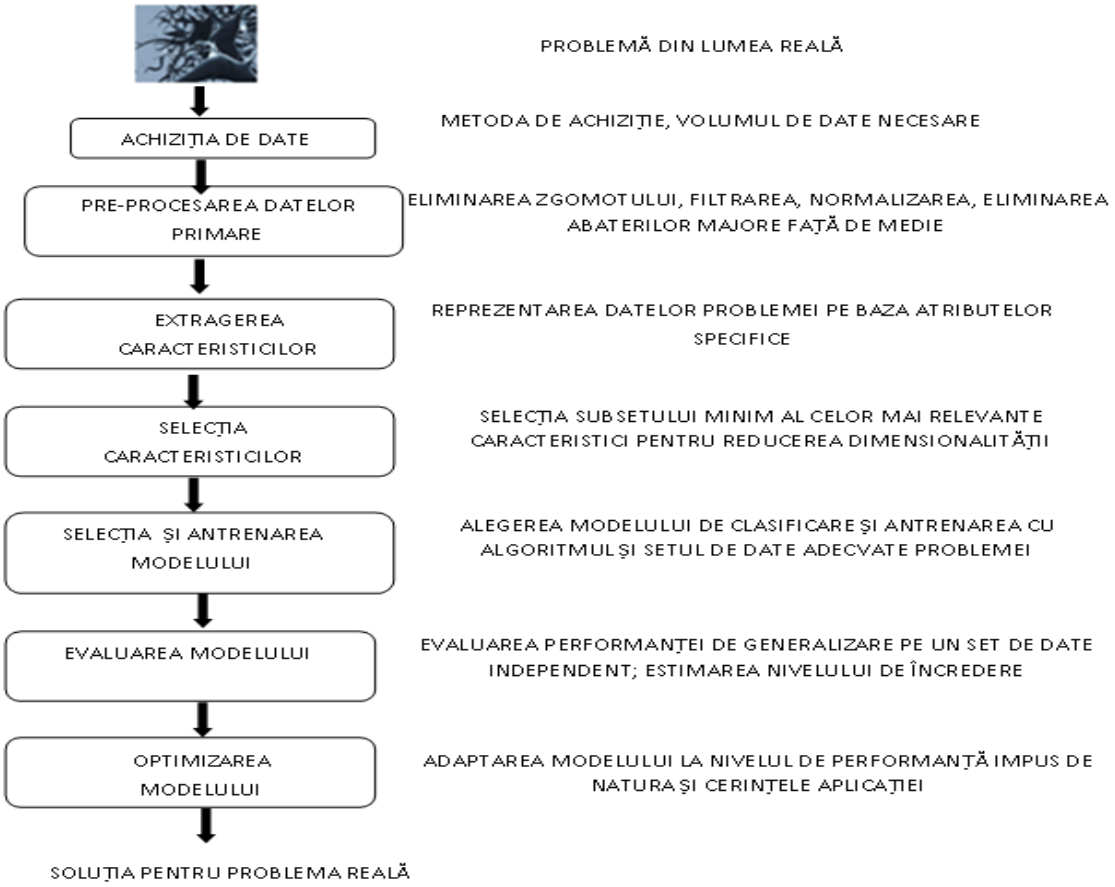

Fig. 1. Componentele unui sistem generic de recunoaştere a paternurilor [3].

Indicatorii de performanță pentru evaluarea sistemelor biometrice (ca sisteme de recunoaştere/ clasificare de paternuri) sunt [5], [6]:

- ***numărul rezultatelor pozitive corecte TP (True Positive)***: numărul de decizii corecte pentru o anumită clasă (identitate) de interes;
- ***numărul rezultatelor fals pozitive FP (False Positive)***: numărul de decizii incorecte pentru clasa (identitatea) de interes;
- ***numărul rezultatelor negative corecte TN (True Negative)***: numărul deciziilor corecte pentru celelalte identități, în afara țintei;
- ***numărul rezultatelor fals negative FN (False Negative)***: numărul deciziilor incorecte pentru apartenența la alte identități excluzând ținta.

Dacă ***Nt*** este numărul total de utilizatori autentici sau numărul de cazuri de identificare a unei persoane, *indiferent de corectitudinea deciziei*, iar ***Nn*** este numărul total de impostori sau, în cazul aplicației de identificare, numărul de cazuri în care o persoană nu este identificată, de asemenea indiferent *de corectitudinea deciziei*, atunci se definesc următorii indicatori relativi ai performanțelor clasificării [4, 5, 6]:

- ***rata rezultatelor corect pozitive TPr***: **senzitivitatea** clasificatorului (*TP ratio*, *recall*):

$$T\Pr = \frac{TP}{Nt} \tag{1}$$

unde numărul total de decizii favorabile pentru identitatea țintă, indiferent de corectitudinea acestora, este

$$Nt = TP + FN \tag{2}$$

- ***rata rezultatelor corect negative TNr***: **specificitatea** clasificatorului (*TN ratio*) :

$$TNr = \frac{TN}{Nn} \tag{3}$$

unde numărul total de decizii favorabile identităților non-țintă, indiferent de corectitudinea acestora, este:

$$Nn = FP + TN \tag{4}$$

- ***precizia clasificatorului*** : capacitatea clasificatorului de a generaliza. Dacă se ține cont de distincția între identitatea de interes și celelalte identități, indicatorul de precizie sau acuratețea clasificării (*Classification Accuracy*) are definiția:

$$CA = \frac{FN}{Nt} + \frac{FP}{Nn} \tag{5}$$

Precizia clasificatorului este evaluată și pe baza raportului dintre numărul de rezultate corect pozitive și numărul total de rezultate pozitive, indiferent de corectitudinea deciziilor:

$$CA = \frac{TP}{TP + FP} \tag{6}$$

Dacă nu se ține cont de distincția între clasa țintă și clasa non-țintă, precizia clasificării este evaluată folosind relația:

$$CA(\%) = \frac{N_c}{N} \cdot 100 = \frac{TP + TN}{Nt + Nn} \cdot 100 \tag{7}$$

unde: N_c este numărul total de șabloane biometrice corect clasificate; N – numărul total de exemple din setul de date considerat.

- ***ratele erorilor de clasificare*** se calculează pe baza numărului de exemple greșit clasificate, pentru fiecare clasă în parte sau prin mediere între toate clasele (rata medie de eroare a clasificatorului, indiferent de clasă). În cazul sistemelor biometrice, semnificațiile indicatorilor de eroare se corelează cu gradul de securitate/insecuritate al sistemului, respectiv cu utilizabilizarea sau rata alarmelor false (capacitatea sistemului de a asigura un grad minim de respingere incorectă a persoanelor autentice). Pentru o clasă dată ω_i, se definesc 2 indicatori de eroare: rata erorilor fals pozitive și rata erorilor fals negative. Definițiile celor 2 indicatori de eroare pentru clasa ω_i sunt următoarele:

  - *rata erorilor fals pozitive pentru clasa* ω_i este

$$\varepsilon_{FP,\omega_j}(\%) = \frac{N_{j,j}}{N_j} \cdot 100 \tag{8}$$

unde: $N_{i,j}$ este numărul exemplelor din clasa reală ω_j ($j \neq i$) clasificate incorect în clasa ω_i, iar N_j este numărul total al exemplelor din clasa reală ω_j, indiferent de corectitudinea deciziei. Dacă sistemul biometric este utilizat pentru **verificarea identității**, iar clasa ω_i este clasa utilizatorilor autentici, atunci acest indicator este o măsură a securității sistemului respectiv, deoarece evaluează probabilitatea de *acceptare incorectă* a unui impostor (**FAR**); în cazul unui sistem biometric cu decizii de **identificare a persoanelor**, rata erorilor fals pozitive pentru o clasă de interes evaluează probabilitatea cu care sistemul proiectat eșuează în recunoașterea persoanei cu identitatea corectă ω_j (eroare de identificare);

- *rata erorilor fals negative pentru clasa* ω_i este

$$\varepsilon_{FP,\omega_i}(\%) = \frac{N_{i,j}}{N_i} \cdot 100 \tag{9}$$

în care: $N_{i,j}$ este numărul total al exemplelor din clasa reală ω_i (clasa pozitivă, clasa utilizatorilor autentici) clasificate incorect în clasa negativă ω_j. Dacă sistemul proiectat este utilizat pentru o aplicație de **verificare** biometrică, iar clasa ω_i desemnează clasa Autentic, atunci acest indicator este o măsură a utilizabilității sistemului biometric, prin evaluarea capacității acestuia de a reduce rata respingerilor false.

În cazul aplicațiilor de **identificare** estimarea celor 2 indicatori ai ratelor de eroare se realizează prin

considerarea unei *clase țintă* (de exemplu, identitatea unei persoane de interes, utilizatorul cel mai privilegiat sau cel mai puțin privilegiat), iar restul șabloanelor biometrice sunt grupate într-o singură clasă non-țintă. O astfel de abordare de tip „unul-vs.-ceilalți" (*one-against-others*) se justifică prin faptul că cei 2 indicatori de eroare evaluează performanța procesului de clasificare numai între perechi de clase.[7]

Toți aceşti indicatori, în particular ratele erorilor de clasificare, stau la baza evaluării preciziei unui sistem biometric folosind indicatorii tipici FAR (FMR) şi FRR (FNMR), prin raportarea acestora la un prag fixat. Analiza de performanță se realizează pe baza curbelor ROC trasate pentru sistemul respectiv. Indicatorii FAR, FMR, FRR, FNMR rezultă din indicatorii de performanță ai unui clasificator, ținând cont de specificul aplicațiilor biometrice.

2.2. Analiza ROC. Principii de aplicare în cazul sistemelor biometrice. Problema separabilității

Indicatorii consacrați pentru caracterizarea preciziei sistemelor biometrice, indiferent de natura aplicației (verificare a identității pretinse sau identificare) se evaluază în raport cu un prag de securitate/ utilizabilitate dependent de cerințele aplicației. În practică, se evaluează *ratele acceptărilor (potrivirilor) false* (False Acceptance Rate **FAR**, respectiv False Matching Rate **FMR**) şi *ratele respingerilor (non-*potrivirilor) false (False Rejection Rate **FRR**, respectiv False Non-Matching Rate **FNMR**):

- ***ratele erorilor de acceptare/respingere falsă*** (**FAR**, **FRR**) se referă la deciziile privind acceptarea sau respingerea unei cereri de acces la resursa protejată. În evaluarea FAR şi FRR se iau în considerare atât cauze interne (legate de imprecizia algoritmilor de extragere a caracteristicilor şi a algoritmilor de clasificare/evaluare a similarității), cât şi externe (legate de condițiile de achiziție a datelor, şi care influențează calitatea şabloanelor biometrice generate);
- ***ratele erorilor de potrivire/non-potrivire falsă*** (**FMR**, **FNMR**) se referă exclusiv la rezultatele returnate de algoritmii de clasificare/evaluare a similarității paternurilor biometrice. În evaluarea acestor rate de eroare se ia în considerare exclusiv imprecizia algoritmilor, nu şi calitatea variată a datelor biometrice de intrare;
- ***rata erorilor egale (EER, Equal Error Rate)***, parametru care caracterizează precizia unui sistem biometric pentru acel prag de securitate (respectiv punct de operare) în care rata erorilor de acceptare (potrivire) falsă este egală cu rata erorilor de respingere (non-potrivire) falsă. Cele 2 tipuri de rate de eroare au evoluții contrare, dar se poate determina (cel puțin teoretic) o valoare a pragului în care cele 2 rate sunt egale sau au valori foarte apropiate.

Perechea de indicatori **(FAR, FRR)**, respectiv **(FMR, FNMR)**, permite evaluarea preciziei unui sistem biometric, prin ***stabilirea pragului optim al deciziilor*** de acceptare/respingere, respectiv al deciziilor de identificare a persoanelor [7]. Evaluarea şi optimizarea performanței unui sistem biometric se realizează prin aplicarea unei tehnici numite **analiză ROC**. **Analiza ROC** constă în totalitatea activităților de modelare şi simulare care permit, în baza unui set consistent de date experimentale, stabilirea unor valori pentru perechi de indicatori de performanță, fiecare pereche de valori fiind obținută pentru un anumit prag fixat prin setările aplicației. Totalitatea acestor perechi de valori formează **curba ROC** pentru sistemul proiectat şi analizat. Un exemplu teoretic de curbă ROC este reprezentat în figura 2 [7]. Deşi o **curbă ROC** este constituită dintr-un număr de puncte separate, fiecare dintre acestea reprezentând un punct de operare al clasificatorului sau al sistemului biometric, prin extrapolare, punctele de operare sunt unite pentru a forma

o reprezentare grafică sugestivă pentru caracterizarea evoluției preciziei sistemului biometric proiectat şi implementat, pentru diferite valori ale pragului de securitate.

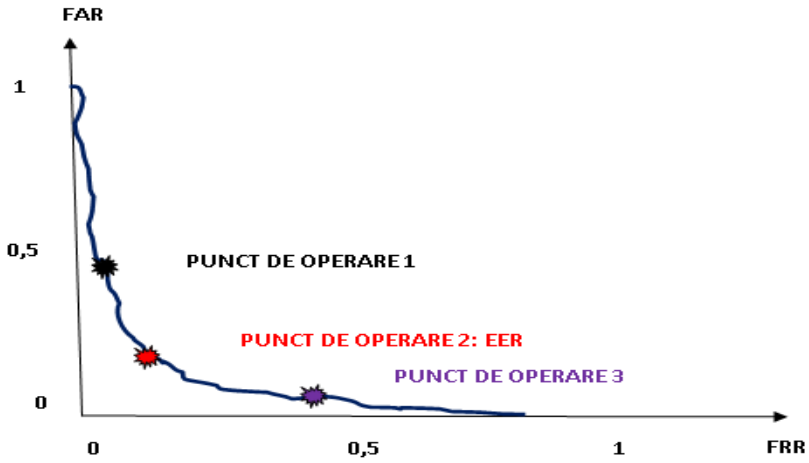

Fig. 2. Exemplu de curbă ROC (3 puncte de operare) [7].

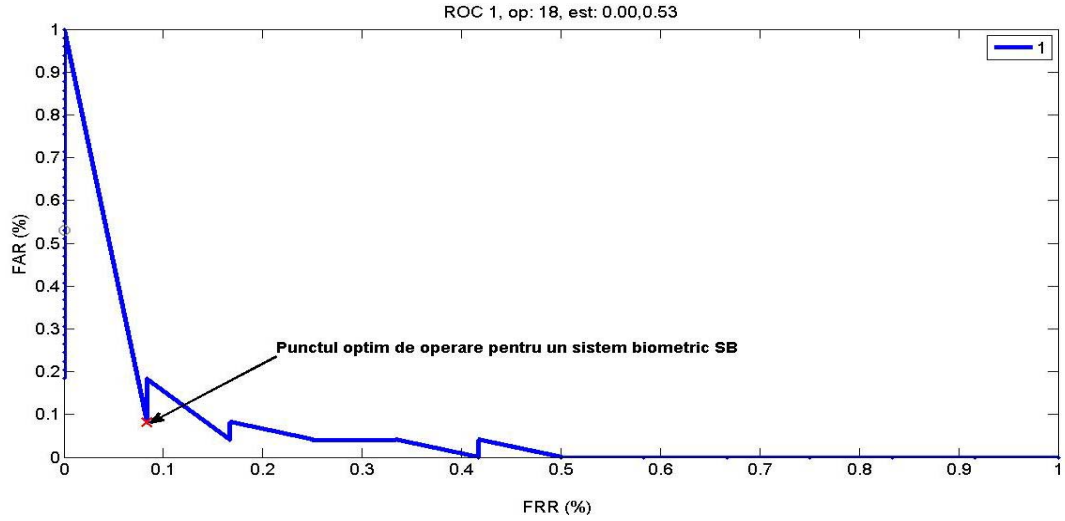

Fig. 3. Curba ROC pentru un sistem biometric unimodal [7].

Analiza performanței unui sistem biometric se bazează pe punctele de operare ale sistemului, stabilite în funcție de pragurile de decizie dependente de aplicație. Astfel, fiecare ***punct de operare*** corespunde unei ***perechi*** (**FAR**, **FRR**) care se obține *pentru o anumită valoare a pragului de decizie de acceptare/respingere* a cererii de acces. **Curba ROC** permite ***determinarea sensibilității sistemului biometric la modificarea pragului de decizie***. Pentru o anumită aplicație, suntem interesați de maximizarea nivelului de securitate, prin urmare punctul de operare al sistemului biometric se deplasează spre dreapta **curbei ROC**, către valori mici ale **FAR**, dar cu dezavantajul creşterii valorilor **FRR** (punctul de operare *3* în figura 2). Pentru o aplicație cu nivel moderat sau redus de securitate, configurarea sistemului biometric se realizează prin fixarea unui punct de operare deplasat spre stânga **curbei ROC**, reducând FRR dar cu creşterea corespunzătoare a FAR (punctul de operare *1* în figura 2). Punctul de operare optimal ar trebui să fie cel care corespunde probabilităților egale de acceptare falsă, respectiv de respingere falsă (punctul de operare *2* în figura 2, care este şi cel mai apropiat de originea sistemului

de axe pentru ambele dimensiuni ale analizei, respectiv ambele tipuri de erori) [7]. Rafinarea analizei **curbei ROC** se realizează prin ***extinderea sau lărgirea scalei de reprezentare a curbelor FAR (FRR) în vecinătatea punctului care corespunde ratei erorilor egale***, aplicând mai multe valori prag în regiunea respectivă a **curbei ROC**. Astfel, se reduce sensibilitatea sistemului la modificările pragului de decizie și crește adaptabilitatea sistemului la cerințele aplicației client. Performanța unui sistem biometric este cu atât mai ridicată cu cât curba sa **ROC** este mai apropiată de axele de coordonate, iar punctul de operare corespunzător ratelor erorilor egale (**EER**) este mai apropiat de originea sistemului de axe. [7, 8, 9]

Analiza posibilităților de optimizare a unui sistem biometric se realizează tot pe baza **curbei ROC** care reprezintă pe axele sistemului de coordonate *cele 2 tipuri de erori de decizie (recunoaştere)* – **FAR (FMR)** şi **FRR (FNMR)** sau, dacă aplicația vizează identificarea persoanelor, *ratele medii de eroare de identificare pentru 2 persoane*. Curba ROC din figura 3 reprezintă evoluția **FAR** în raport cu **FRR** pentru un sistem biometric bazat pe amprentă (caz real, curbă obținută din date experimentale).

Curba ROC permite compararea obiectivă a mai multor sisteme biometrice. Analiza unei **curbe ROC**, în absența specificării unor criterii suplimentare, arată că un sistem biometric poate fi mai bun ***la nivel global*** (*pentru toate punctele de operare admisibile*), dar la nivel local (*pentru un set restrâns de puncte de operare localizate în vecinătatea unui anumit punct de interes*), comportamentul sistemului poate prezenta abateri față de tendința globală (comparativ cu alte sisteme biometrice).

Pentru a lua în considerare şi aceste situații, se utilizează un criteriu numit **separabilitate**. **Separabilitatea** unui sistem biometric este capabilitatea acestuia de a distinge cu precizie maximă între utilizatorii autorizați şi cei neautorizați, pe baza unei anumite trăsături fizice sau comportamentale aplicate la intrarea sistemului. Separabilitatea este cu atât mai ridicată cu cât ratele erorilor de decizie sunt mai mici. În principiu, *punctul optim de operare al sistemului biometric*, care corespunde egalității ratelor de eroare (**EER**), şi care este cel mai apropiat de originea sistemului de axe pe ambele sale dimensiuni (figura 3) este şi cel care maximizează separabilitatea sistemului biometric.

Orice criteriu de evaluare a separabilității unui sistem biometric trebuie să fie independent de gama de valori ale indicatorilor de performanță utilizați (ratele erorilor de decizie sau de recunoaştere), şi trebuie să fie uşor de evaluat. De exemplu, ***rata erorilor egale*** (**EER**) este un indicator utilizabil pentru cuantificarea separabilității. **EER** este asociat unui singur punct de operare de pe **curba ROC** a sistemului biometric, dar nu există ca valoare direct măsurabilă, ci se obține prin decizie şi aproximare, folosind valorile celor 2 indicatori ai ratelor erorilor de decizie. [10], [11]

O altă măsură a separabilității este ***aria de sub curba ROC***, care se obține prin însumarea valorilor ROC pentru toate punctele de operare. Fiecare *punct al curbei ROC* (*punct de operare* al sistemului biometric) corespunde unui *anumit prag de decizie de acceptare/respingere, respectiv de identificare a persoanelor*, prag fixat pentru sistem, fie prin proiectare, fie la punerea în funcțiune, în acord cu cerințele aplicației. Problema principală în cazul utilizării acestui criteriu este că, de regulă, valorile ROC nu sunt echidistante, deoarece multe aplicații nu sunt configurate pentru praguri de decizie distribuite în paşi egali [7], [11] (figura 4).

De exemplu, în curba ROC a unui sistem biometric reprezentată în figura 4, zona de securitate maximă (valori minime ale **FAR** şi valori mari ale **FRR**) prezintă o densitate mai mare de puncte de operare,

deoarece aplicația client a impus ajustarea optimă a nivelului de securitate, prin fixarea unui număr mai mare de praguri de decizie. În plus, analiza ROC pe un număr mai mare de puncte de operare permite studiul detaliat al efectelor diferitelor configurări ale aplicației client asupra ratelor erorilor de decizie. În aceste condiții, uniformitatea analizei se asigură de exemplu prin ponderarea fiecărei valori **FAR** (reprezentate pe axa *y*) prin *distanța dintre valorile succesive* reprezentate pe axa x (valorile **FRR** pentru 2 puncte de operare consecutive fixate). Această distanță se obține folosind reprezentarea grafică a funcțiilor distribuție de probabilitate pentru scorurile utilizatorilor neautorizați . Dacă se admite că funcția de repartiție se aplică pentru o variabilă aleatoare continuă, suma punctelor succesive devine integrală pe domeniul care corespunde scorurilor impostorilor [11].

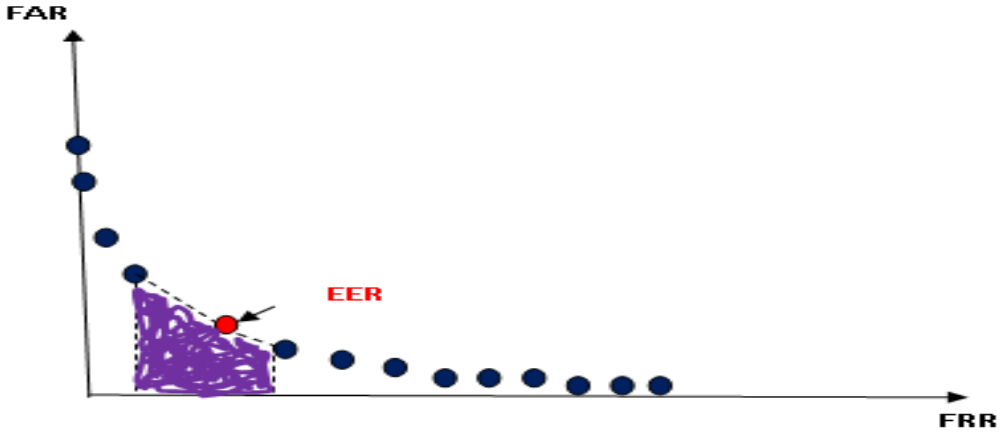

Fig. 4. Puncte de operare neechidistante pentru un sistem biometric (praguri de decizie inegal distanțate) [7].

Pentru un număr *N* de scoruri de similaritate, aria ROC se obține astfel:

$$A_{ROC} = \sum_{s_i, i=1}^{N} FRR(s_i) \cdot p_I(s_{i-1}) \qquad (10)$$

în care p_I este funcția distribuție de probabilitate pentru scorurile impostorilor. Relația (10) este o sumă de valori dependente de scorurile calculate de algoritmul de recunoaştere. Condiția de independență a ariei ROC A_{ROC} de definiția analitică a scorului de similaritate impune absența respingerilor independente de prag (**FRR** = **FNMR**), caz în care în evaluarea ariei ROC se utilizează doar rata erorilor de nepotrivire falsă, estimată ca funcție de distribuție cumulată. Se ia în considerare numai numărul de scoruri de similaritate dintr-un anumit domeniu de interes (de exemplu zona utilizatorilor impostori), respectiv frecvența relativă a acestora, nu şi modul de calcul al acestora.[11]

Un sistem biometric asigură o **separabilitate ideală** între utilizatorii autorizați şi neautorizați (în condițiile disponibilității unor estimatori relevanți ai funcțiilor de distribuție pentru scorurile utilizatorilor autentici şi impostori, p_A şi, respectiv p_I) dacă: $EER = 0$ şi $A_{ROC} = 0$. Condiția de **separabilitate ideală** impune inexistența nici unei suprapuneri între cele 2 distribuții p_A şi p_I. Nici un sistem biometric real nu îndeplineşte această condiție, suprapunerile dintre distribuții fiind determinate de factori externi dar şi de factori specifici algoritmilor de recunoaştere.

3. ANALIZĂ DE CAZ PRIVIND OPTIMIZAREA PERFORMANȚELOR UNUI SISTEM BIOMETRIC PRIN SELECȚIA ADECVATĂ A PUNCTELOR DE OPERARE

Studiul de caz pentru exemplificarea aplicării analizei ROC în vederea evaluării şi optimizării performanțelor unui sistem biometric include:

- prezentarea arhitecturii de sistem biometric multimodal;
- explicitatea modulului de calcul al scorurilor de similaritate;
- prezentarea regulii de fuziune post-clasificare la nivel de scor de similaritate;
- realizarea analizei ROC şi optimizarea performanțelor sistemului.

3.1. Arhitectura de sistem

Considerăm un sistem biometric multimodal cu 3 componente de identificare: un subsistem de identificare pe baza amprentelor, un subsistem de identificare pe baza modelului palmar şi un subsistem de identificare bazat pe iris. Arhitectura de sistem este reprezentată în figura 5. Pentru combinarea scorurilor de similaritate se aplică o regulă de fuziune post-clasificare la nivel de scor, regulă bazată pe suma ponderată a scorurilor de similaritate calculate pentru fiecare tip de şablon biometric.

3.2. Calculul scorurilor de similaritate. Tehnici de normalizare aplicate

Fie *S* ***scorul de similaritate*** între vectorul de caracteristici x şi şablonul biometric corespunzător (datele de referință) *z*: $S = s(x,z)$. Deoarece comparația dintre vectorul de caracteristici şi şablonul corespunzător se realizează prin calculul unei distanțe în spațiul caracteristicilor, procedura de calcul al scorului de similaritate include următoarele etape:

- *calculul distanței d(x, z);*
- *normalizarea scorului de tip distanță;*
- *transformarea scorului de tip distanță în scor de similaritate.*

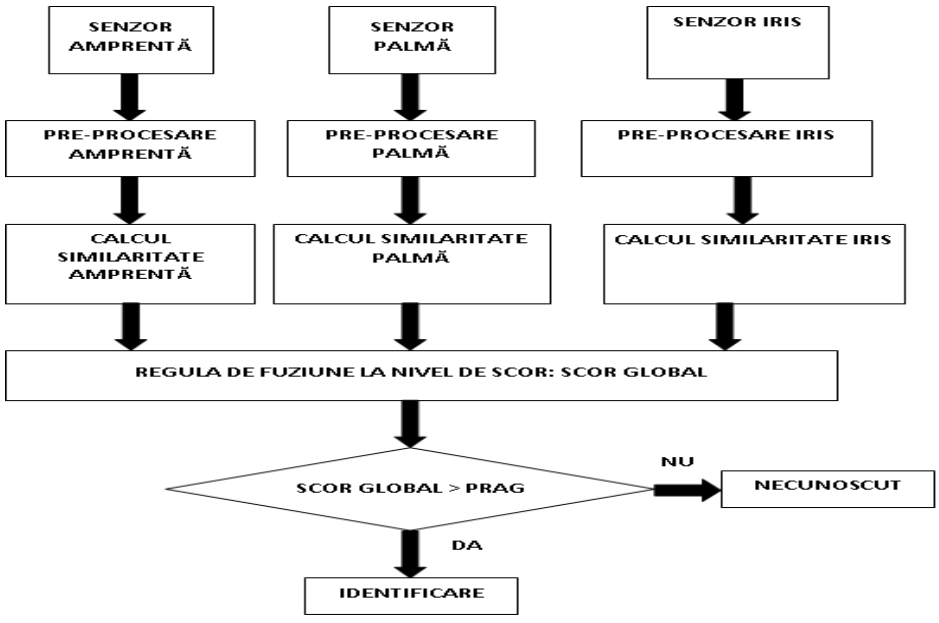

Fig. 5. Structura sistemului de securitate bazat pe integrarea a 3 tehnologii biometrice.

3.2.1. Calculul distanței între vectorul curent şi şablonul biometric de referință

Cuantificarea similarității dintre vectorul curent de caracteristici *x* şi vectorul de referință *z* se realizează utilizând o **metrică de distanță în spațiul caracteristicilor**, pentru fiecare tip de date biometrice. Pentru calculul scorurilor de similaritate am aplicat **distanța Mahalanobis**, care este o distanță bazată pe corelația dintre variabile, utilizabilă pentru determinarea similarității dintre 2 seturi de caracteristici. Distanța Mahalanobis dintre 2 vectori *u* şi *v*, de dimensiuni egale *d*, este dată de relația

$$d_M(u,v) = \sqrt{(u-v)^T \cdot \Sigma^{-1} \cdot (u-v)} \qquad (11)$$

în care ∑ este matricea de covarianță a celor 2 vectori.

Alegerea acestui tip de distanță se justifică prin proprietățile acesteia: ***exploatarea corelației dintre componentele vectorilor de caracteristici*** şi ***invarianța la scalare***.

3.2.2. Normalizarea scorului de tip distanță

Prin **normalizarea scorurilor** se realizează transformarea scorurilor individuale pentru a fi aduse la un domeniu comun de valori numerice, înainte de aplicarea schemei de fuziune a acestora. Pentru modelul propus de sistem biometric multimodal (cu 3 componente de identificare), **tehnica de normalizare aplicată** se bazează pe ***funcția sigmoidă***:

$$f(x) = \frac{1}{1+\exp(-x)}, f: \mathbb{R} \to [0,1] \qquad (12)$$

Pentru orice valoare reală din domeniul de definiție, valoarea funcției aparține intervalului [0,1].

Normalizarea scorurilor de similaritate se va realiza folosind *forma simplă a funcției sigmoide*, dar şi *funcția dublu sigmoidă*.

Forma normalizată a scorului de tip ***distanță*** pentru vectorul de caracteristici x_i, folosind *forma simplă a funcției sigmoide*, este:

$$D_i(x_i, z_{\hat{i}}) = \frac{1}{1+A_i \cdot \exp(-B_i \cdot d_{M,i}(x_i, z_i))}, \quad i = \overline{1,3} \qquad (13)$$

relație în care coeficienții A_i şi B_i sunt stabiliți pentru fiecare tip de vector de caracteristici biometrice, utilizând seturile de date de referință disponibile.

Dacă se utilizează *funcția dublu sigmoidă* pentru normalizarea scorului de ***distanță***, expresia acestuia devine

$$D_i(x_i, z_{\hat{i}}) = \begin{cases} \dfrac{1}{1+A_i \cdot \exp\left[-B_i \cdot \left(\dfrac{d_{M,i}(x_i,z_i)-\theta}{C_{1,i}}\right)\right]}, & \text{pentru } d_{M,i}(x_i,z_i) < \theta \\ \dfrac{1}{1+A_i \cdot \exp\left[-B_i \cdot \left(\dfrac{d_{M,i}(x_i,z_i)-\theta}{C_{2,i}}\right)\right]}, & \text{pentru } d_{M,i}(x_i,z_i) \geq \theta \end{cases} \qquad (14)$$

în care:

- coeficienții A_i şi B_i ,stabiliți pe baza datelor experimentale, reprezintă parametri de formă ai funcției sigmoide: pentru *amprentă* $A_1 = 1,5$ şi $B_1 = 2$; pentru *palmă* $A_2 = 2,5$ şi $B_2 = 1$; pentru *iris* $A_3 = 1$ şi $B_3 = 2,25$;
- $C_{1,i}$ şi $C_{2,i}$ reprezintă marginile regiunii în care funcția sigmoidă este cvasi-lineară: pentru *amprentă* $C_{1,1} = 1,5$ şi $C_{2,1} = 1,75$; pentru *palmă* $C_{1,2} = 2,5$ şi $C_{2,2} = 3$; pentru *iris* $C_{1,3} = 1,25$ şi $C_{2,3} = 4$;
- θ este o valoare prag corelată cu nivelul de securitate/acceptabilitate al sistemului biometric.

3.2.3. Transformarea scorului de tip distanță în scor de similaritate

Scorul normalizat de tip distanță $D(x_i, z_{\hat{i}})$ cuantifică diferența dintre vectorul de caracteristici x_i şi şablonul biometric z_i (de referință). Cu cât valoarea *scorului de tip distanță* este mai mare, cu atât diferența dintre vectorii comparați este mai mare. Spre deosebire de scorul de tip distanță, *scorul de similaritate* cuantifică gradul de similitudine dintre vectorii de caracteristici. Cu cât valoarea scorului de similaritate este mai ridicată, cu atât numărul de caracteristici similare creşte, astfel încât probabilitatea de autenticitate a utilizatorului este mai mare.

Deoarece domeniul de valori ale scorurilor bazate pe distanțe normalizate prin aplicarea funcției sigmoide este intervalul [0,1], transformarea *scorului de distanță* în *scor de similaritate* se realizează astfel:

$$S_i = s(x_i, z_i) = 1 - D_i(x_i, z_i),\ i = \overline{1,3} \qquad (15)$$

3.3. Regula de fuziune post-clasificare la nivel de scor de similaritate

Pentru fuziunea post-clasificare (la nivel de scor) am aplicat **regula sumei ponderate a scorurilor** obținute pentru fiecare tip de date biometrice. Ponderile sunt selectate în funcție de performanța fiecăruia dintre cele 3 subsisteme de identificare (amprentă, model palmar şi iris).

3.4. Analiza ROC. Determinarea punctelor optime de operare

Analiza ROC în cazul sistemului propus constă în ***stabilirea punctelor de operare*** pentru fiecare subsistem în parte şi pentru sistemul multimodal în ansamblu. Se au în vedere cele 2 variante de normalizare a scorurilor. **Optimizarea performanței** sistemului biometric multimodal constă în ***adaptarea acestuia la cerințele de securitate/utilizabilitate ale aplicației***, prin ***selecția/fixarea acelor puncte de operare*** care asigură un *optim între* **securitate** (prin *reducerea ratei acceptărilor false*) *şi* **utilizabilitate/acceptabilitate** (prin *reducerea ratei respingerilor false*). Deoarece evoluția valorilor celor 2 indicatori ai preciziei unui sistem biometric este opusă, în mod tipic punctul optim de operare corespunde acelei valori a pragului de securitate care asigură egalitatea ratelor erorilor de acceptare falsă şi de respingere falsă. Deoarece în cazul sistemului propus nu am obținut nici un punct de operare care să corespundă egalității FAR şi FRR, am selectat ca punct optim de operare, în fiecare caz, acel punct al curbei ROC aflat cel mai aproape de originea sistemului de axe de coordonate.

Figura 6*a,b* reprezintă grafic curbele ROC pentru sistemul biometric multimodal şi pentru cele 3 subsisteme de identificare integrate (amprentă, palmă, iris), în condițiile în care normalizarea scorurilor de similaritate se realizează folosind funcția sigmoidă simplă (figura 6*a*), respectiv funcția dublu sigmoidă (figura 6*b*). Punctele optime de operare pentru sistem şi pentru subsistemele componente sunt caracterizate prin perechile de indicatori FAR-FRR ale căror valori sunt date în tabelul 1. Fiecare pereche de valori FAR-FRR defineşte un punct de operare obținut pentru un anumit prag de securitate fixat la nivel de aplicație.

În toate cazurile, punctele optime de operare sunt caracterizate de valori mai mici ale FRR comparativ cu valorile FRR. Am admis un compromis între gradul de securitate şi cel de acceptabilitate, pentru a reduce probabilitatea de respingere falsă sub 0,4%. Pe de altă parte, aceste rezultate, obținute din datele experimentale disponibile, confirmă faptul că punctul de operare corespunzător EER este unul

teoretic, greu de atins în practică, deşi indicatorul EER este utilizat de mulți furnizori de senzori biometrici şi integratori de sisteme biometrice ca măsură a preciziei sistemului proiectat pentru aplicațiile client.

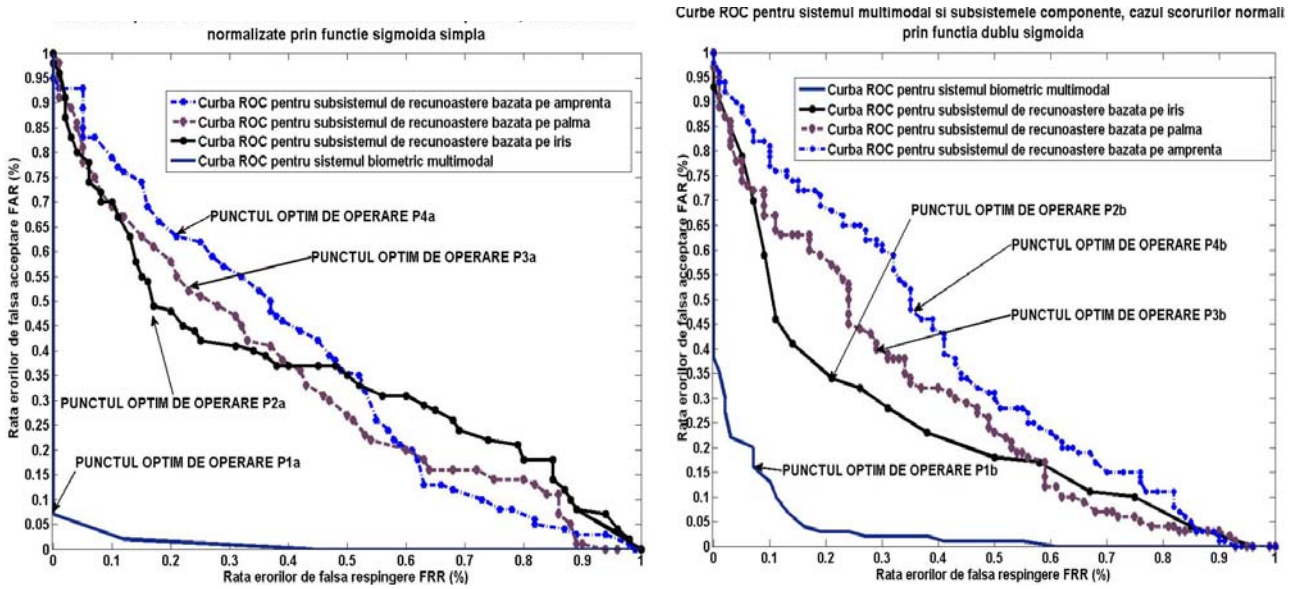

Fig. 6. Curbe ROC pentru sistemul biometric multimodal şi subsistemele componente:
a – cazul normalizării prin funcție sigmoidă simplă; *b* – cazul normalizării prin funcție sigmoidă dublă.

Tabelul 1

Punctele optime de operare rezultate din aplicarea analizei ROC

Punctul optim de operare	FAR (%)	FRR (%)
Cazul a) Scoruri normalizate prin funcția sigmoidă simplă		
P1a: *sistemul multimodal*	0,075	0,001
P2a: *subsistemul iris*	0,49	0,16
P3a: *subsistemul palmă*	0,525	0,225
P4a: *subsistemul amprentă*	0,630	0,210
Cazul b) Scoruri normalizate prin funcția sigmoidă dublă		
P1b: *sistemul multimodal*	0,155	0,055
P2b: *subsistemul iris*	0,34	0,205
P3b: *subsistemul palmă*	0,4	0,275
P4b: *subsistemul amprentă*	0,475	0,35

Pe de altă parte, este de observat că, pentru datele disponibile, normalizarea scorurilor de similaritate prin funcție sigmoidă dublă nu îmbunătățeşte performanțele sistemului. În ambele cazuri (funcție sigmoidă simplă şi dublă), performanța sistemului multimodal se îmbunătățeşte semnificativ comparativ cu performanțele subsistemelor individuale de identificare.

4. CONCLUZII

Proiectarea şi implementarea de sisteme de securitate integrând factori biometrici multipli este o preocupare actuală majoră în domeniul soluțiilor de securitate pentru diferite clase de resurse accesibile folosind rețele de comunicații publice şi cu grad scăzut de protecție (în principal Internetul). În acest context, abordarea propusă, de analiză şi optimizare a performanțelor unui sistem biometric multimodal prin selecția adecvată a punctelor de operare pe curbele ROC, asigură un cadru facil şi intuitiv pentru proiectarea şi evaluarea de soluții biometrice de securitate; în acest caz, optimizarea se referă la capacitatea soluției de adaptare la cerințele specifice ale aplicației, deoarece nu toate aplicațiile practice ale sistemelor biometrice impun aceleaşi condiții

privind pragul de securitate, utilizabilitatea şi costurile de implementare.

Sistemul proiectat şi optimizat se bazează pe fuziunea post-clasificare la nivel de scor, care este uşor de implementat şi nu este dependentă de formatul şabloanelor biometrice (care în mod tipic este puțin acccesibil unui integrator de sistem). Totuşi, una dintre direcțiile actuale de cercetare în domeniul biometriei vizează fuziunea la nivel de caracteristici (fuziunea pre-clasificare), deoarece combinarea mai multor caracteristici biometrice independente are un potențial semnificativ de creştere a preciziei identificării persoanelor. Aceasta impune însă considerarea unor seturi extinse de date, cu diferite tehnici de extragere a caracteristicilor, ținând însă cont şi de compatibilitatea dintre caracteristicile rezultate.

Bibliografie

[1] **Soviany Sorin, Puşcoci Sorin, Jurian Mariana:** *A Hierarchical Data Classification Model for Biometric Identification Systems*, The 3rd International Conference on Emerging Intelligent Data and Web Technologies (EIDWT 2012), Universitatea „Politehnica“, Bucureşti, 19-21 septembrie 2012

[2] **Soviany Sorin, Jurian Mariana, Puşcoci Sorin:** *Securizarea accesului la sisteme informatice prin metode biometrice multimodale*, Sesiunea de comunicări ştiințifice *„Rolul şi implicarea cercetării ştiințifice în dezvoltarea şi implementarea operațională a echipamentelor şi sistemelor pentru securitate şi apărare”,* Agenția de Cercetare pentru Tehnică şi Tehnologii Militare, Bucureşti, 29 noiembrie 2010

[3] **Robi Polikar:** *Pattern recognition*, Wiley Encyclopedia of BioMedical Engineering, 2006

[4] **Webb Andrew R., Copsey Keith D.:** *Statistical Pattern Recognition*, 3rd edition, Wiley, 2011

[5] ***: Curs *Pattern Recognition: Classification, Discriminant Analysis*, Universitatea Delft, Olanda, 2009-2010

[6] ***: *PerClass Training Course: Machine Learning for R&D Specialists*, Delft, Olanda

[7] **Soviany Sorin:** Teză de doctorat *Optimizarea deciziei în sistemele de identificare biometrică*, Universitatea Piteşti, ianuarie 2013

[8] ***: *Study Report on Biometrics in E-Authentication,* M1.4 Ad-Hoc Group on Biometric in E-Authentication, International Committee for Information Technology Standards (INCITS), INCITS M1/07-018rev, 30 martie 2007

[9] **Jain A., Ross A., Prabhakar S.:** *An Introduction to Biometric Recognition*, IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, Nr. 1, 2004

[10] ***: *Biometric Technology Application Manual*, Volume 1: *Biometrics Basics*, *National Biometric Security Project, 2005-2007*

[11] **Bromba Manfred:** *Bioidentification. Frequently Asked Questions. Biometrics*, FAQ adopted to the ISO/IEC Harmonized Biometric Vocabulary.