

# *Capitolul 1*

## **Aplicații biometrice. Recunoașterea fețelor**

- 1.1 Recunoașterea fețelor
- 1.2 Algoritmi de recunoaștere a fețelor
- 1.3 Aplicații ale algoritmilor de  
recunoaștere a fețelor
- 1.4 Discuție asupra sistemelor biometrice

Identitatea unei persoane este reprezentată nu de actele oficiale pe care le deține, ci de un set extins, complex, variabil în timp și nu întotdeauna ușor de definit de trăsături personalizate, de natură anatomică, fiziologică sau comportamentală. Unele dintre aceste trăsături (considerate izolat sau sub forma unor combinații) pot fi folosite, împreună cu tehnici automate de procesare a semnalelor, pentru implementarea așa-numitelor *sisteme biometrice*, capabile să recunoască sau să valideze autenticitatea identității unor persoane. Drept informații biometrice se utilizează în mod curent amprente, vocea, fața, irisul, forma geometrică a mâinii. Spre deosebire de modalitățile tipice de identificare precum legitimațiile, cheile, parolele sau codurile PIN (*Personal Identification Number*), informațiile biometrice nu pot fi pierdute, furate sau uitate (însă uneori pot fi reproduse cu acuratețe suficientă pentru a “păcăli” sistemele automate de recunoaștere).

Pentru definirea și măsurarea performanțelor trebuie avute în vedere următoarele elemente ce compun un *sistem biometric ideal*:

- toți membrii populației posedă caracteristicile pe care sistemul le identifică
- fiecare “semnătură” biometrică a unei persoane diferă de semnăturile tuturor celorlalte persoane din populația supusă analizei
- “semnătura” biometrică nu variază semnificativ în funcție de condițiile particulare în care este extrasă (obținută)
- sistemul rezistă tentativelor de fraudare a semnăturilor

Metodologiile de evaluare a sistemelor biometrice cuantifică măsura în care sistemele reale se apropie de aceste cerințe ideale. Sunt avute în vedere două clase de aplicații, *de identificare (recunoaștere)*, respectiv *de verificare (autenticare)*:

- într-un sistem de identificare (recunoaștere), “semnătura” biometrică a unei persoane necunoscute este prezentată la intrarea sistemului, urmând a fi comparată cu cele disponibile într-o bază de date alcătuită din “semnături” ale unui set de persoane cunoscute. Sistemul furnizează ca răspuns identitatea persoanei din baza de date a cărei “semnătură” seamănă cel mai bine – conform unei distanțe (metrici) definite adecvat! – cu cea a persoanei necunoscute (în principiu există și posibilitatea ca sistemul să decidă că persoana necunoscută nu seamănă cu nimeni din baza de date, dacă distanța minimă depășește totuși o anumită valoare de prag).

Ca exemple amintim situațiile în care poliția urmărește identificarea unui infractor pe bază de amprente sau fotografii, precum și situațiile în care aceeași persoană încearcă să depună în mod fraudulos mai multe documentații în vederea obținerii unor avantaje materiale (de exemplu, depunerea a mai multor dosare de obținere a pașaportului sau a carnetului de conducere).

- într-o aplicație de verificare, o persoană prezintă “semnătura” sa biometrică și pretinde că are identitatea care corespunde acelei “semnături”, pe baza căreia vrea să obțină dreptul de acces într-un spațiu sau la anumite resurse restricționate. Sistemul poate accepta sau respinge această cerere, respectiv poate furniza un grad de încredere asupra validității identității pretinse. Ca exemple practice putem menționa autentificarea identității în tranzacțiile bazate pe utilizarea cardurilor, precum și accesul restricționat la anumite calculatoare sau incinte de clădiri.

Este important de subliniat că indicatorii de performanță utilizați în sistemele de verificare diferă net de cei utilizați de sistemele de identificare (recunoaștere). Astfel, calitatea unui sistem de identificare este măsurată de procentajul de răspunsuri corecte dintr-un număr suficient de mare de teste. De exemplu, pentru identificarea unui suspect, poliția utilizează o bază de date în format electronic ce conține fotografii pentru un număr foarte mare de infractori. Un test tipic constă în furnizarea la intrarea sistemului a fotografiei unei persoane necunoscute, iar răspunsul sistemului constă într-o listă sortată de fotografii (sortată de regulă în ordine descrescătoare a gradului de similaritate cu imaginea persoanei necunoscute). Astfel, indicatorul de performanță avut în vedere este procentajul de situații în care identitatea persoanei necunoscute se regăsește cu adevărat printre primele “k” răspunsuri ale sistemului (tipic,  $k=5$ ). Pe de altă parte, în cazul sistemelor de verificare, se utilizează în mod tradițional două tipuri de erori:

- rată falsă de rejecție (*False Reject Rate*)
- rată falsă de acceptare (*False Acceptance Rate*)

O rejecție falsă apare când sistemul refuză o identitate validă, iar o acceptare falsă apare atunci când sistemul acceptă o identitate falsă. Într-un sistem biometric ideal, ambele rate de erori ar fi zero. În practică, sistemele realizează un anumit compromis. De regulă, sistemele se proiectează pentru a avea o valoare fixată a

ratei de acceptare falsă, ce va conduce în mod corespunzător la determinarea valorii ratei de rejecție falsă. Alegerea acestor parametri depinde de aplicația avută în vedere. Astfel, la cititoarele de carduri (ATM), rata de rejecție falsă va fi aleasă mică pentru a nu crea disconfort clienților de bună credință. Pe de altă parte, în situațiile de acces restricționat în anumite spații, factorul principal avut în vedere va fi alegerea unei rate false de acceptare cât mai mici. Recent a fost propusă o generalizare a aplicațiilor de verificare denumită generic *watch list*, în care verificarea identității unei persoane necunoscute se efectuează în raport cu o listă de persoane “suspecte” și nu față de o singură persoană.

Arhitectura unui sistem biometric generic include următoarele module [18]:

- a) un senzor utilizat pentru colectarea informațiilor primare și conversia acestora în format digital
- b) un algoritm de procesare de semnal care va extrage o „semnătură” biometrică adecvată
- c) o bază de date în care se stochează „semnături” provenind de la o populație de subiecți
- d) o procedură de comparație a “semnăturii” corespunzătoare unei persoane necunoscute cu cele stocate în baza de date
- e) o procedură de decizie (complet automată sau asistată de către om) care utilizează rezultatul comparației anterioare în vederea efectuării unei acțiuni.

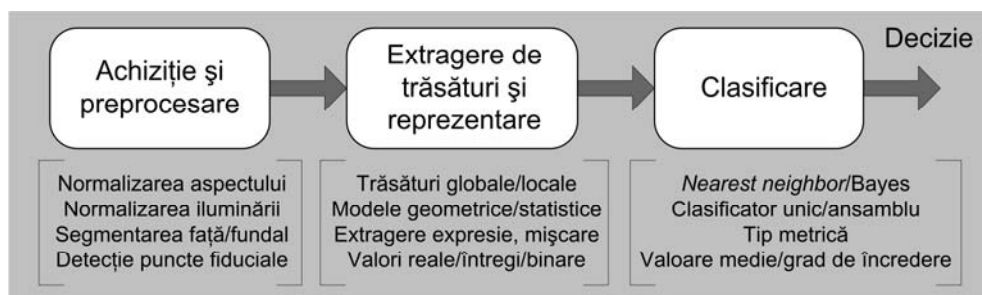
Este important de subliniat că datorită dificultății aplicațiilor biometrice, precum și a cerințelor de performanță impuse, se folosesc deseori sisteme complexe, care urmăresc recunoașterea/identificarea unei persoane folosind mai multe tipuri de “semnături”, de exemplu imaginea feței (sau amprenta) și vocea. Pentru testarea sistemelor biometrice au fost elaborate protocoale de evaluare standardizate. Astfel, pentru recunoașterea fețelor, institutul american NIST (*National Institute of Standards and Technology*) a elaborat standardul FERET (*Face Recognition Technology*) și a efectuat de asemenea evaluări ale tehnologiei de recunoaștere a vorbitorului. De regulă, astfel de evaluări fac obiectul unor competiții deschise atât companiilor cât și universităților și sunt coordonate de către un grup de experți independenți.

## 1.1 Recunoașterea fețelor

Fața joacă un rol esențial în relațiile sociale, în comunicarea identității și a emoțiilor. Capacitatea umana de a recunoaște fețele este remarcabilă: putem recunoaște mii de figuri învățate de-a lungul vieții și identifica fețele familiare dintr-o singură privire chiar după ani de zile. Această capacitate este destul de robustă, în ciuda schimbărilor mari în stimulul vizual datorate condițiilor de vizualizare, expresiei, îmbătrânirii și altor factori perturbatori precum prezența ochelarilor, a bărbii sau schimbări ale coafurii. Ca urmare, acest subiect a fost în centrul atenției unor grupuri numeroase de cercetători din întreaga lume în ultimele două decenii și a pătruns de curând și în circuitul comercial: se estimează că piața aplicațiilor biometrice va depăși 2 miliarde \$ până la sfârșitul lui 2006, ajungând la aproximativ 5.5 miliarde \$ în 2010, cu o rată de creștere anuală de aproximativ 25% [8]. Pe de altă parte, în ciuda numeroaselor soluții originale raportate în literatură, tehnicile actuale sunt încă departe de a oferi nivele de performanță acceptabile pentru introducerea acestor tehnologii în domenii sensibile precum controlul identității călătorilor în aeroporturi sau controalele vamale.

Deși atractivă deoarece nu presupune contact direct între subiect și senzor (se spune că este o tehnică *neinvazivă*), recunoașterea feței este o sarcină dificilă datorită surselor numeroase de variabilitate în condiții reale. Acestea includ printre altele: orientarea feței în raport cu aparatul de fotografiat sau camera de filmat, nivelul de iluminare, expresia feței, timpul scurs între momentele prelevării unor imagini distincte, precum și aspectele demografice (rasa, vârsta, sexul). Asigurarea invarianței în raport cu transformări elementare precum translația, rotația sau schimbarea rezoluției reprezintă de asemenea o cerință obligatorie. Schema-bloc a unui sistem generic de recunoaștere a fețelor se prezintă în Fig. 1.1, în care se distinge prezența unui modul de extragere a “semnăturii” specifice fiecărei fețe și a unui modul de clasificare, a căror proiectare este de regulă corelată.

Există 2 abordări majore pentru a obține informația caracteristică unei fețe, cu avantaje și dezavantaje specifice, ce pot fi folosite pentru extragerea “semnăturilor” de interes atât pentru aplicații de recunoaștere cât și pentru cele de verificare, diferențe apărând datorită tipului de clasificator utilizat:



**Fig. 1.1** Schema-bloc a unui sistem de recunoaștere/verificare a fețelor

a) algoritmi bazați pe analiza statistică a imaginilor disponibile, al căror scop constă în identificarea unei baze reprezentative în raport cu care să poată fi exprimată orice imagine sub forma unei combinații liniare de vectori ai bazei. Din această categorie fac parte metode precum Analiza pe Componente Principale (PCA) [24], Analiza Discriminatorie Liniară (LDA) [2] și Analiza pe Componente Independente (ICA) [1].

b) algoritmi bazați pe măsurarea unor trăsături geometrice referitoare la distanțe între puncte semnificative de pe suprafața feței. Ca exemple putem enumera metodele *Elastic Bunch Graph Matching* [25] și *Local Feature Analysis* [19].

Câteva studii recente [9, 28] trec în revistă tendințele actuale din acest domeniu, printre care menționăm:

- extensii de tip nucleu (*kernel*) ale unor metode clasice de proiecție pe subspații liniare, care au condus la apariția algoritmilor de tip *Kernel PCA/LDA/ICA*
- algoritmi bazați pe utilizarea unor reprezentări folosind baze cu bună localizare spațială, respectiv metode urmărind descompunerea fețelor în componente distincte (gupate în jurul unor puncte reprezentative precum ochii, nasul, gura – denumite generic *puncte fiduciale*)
- analiza imaginilor 3D și aplicarea tehnicilor de *morfism* (care urmăresc generarea unor *imagini virtuale* ale unei persoane pornind de la un set de imagini captate din unghiuri neconvenabile)
- un număr semnificativ de articole recente au ca element comun abordarea bazată pe modelarea și aproximarea distribuției spațiale a imaginilor reprezentând fețe umane sub forma așa-numitelor varietăți matematice (*manifolds*).

Alte aspecte interesante se referă la asigurarea unei toleranțe sporite în raport cu numeroasele surse de variabilitate specifice acestor aplicații (grad de iluminare variabil, poziție, transformări geometrice, acoperire parțială), elaborarea unor metode capabile să ofere performanțe rezonabile în condițiile existenței unor baze de date de dimensiune redusă, detecția automată a punctelor fiduciale.

Înțelegerea profundă a conținutului informației despre imaginile care reprezintă fețe umane este o cerință fundamentală pentru succesul oricărui sistem performant de recunoaștere sau verificare și se poate dovedi folositor și în alte aplicații de recunoaștere a formelor (*pattern recognition*). Enumerăm în continuare principalele aspecte critice referitoare la acest subiect, indicate în Fig. 1.2, prezentându-le ca pe un sumar de întrebări fundamentale și indicând soluțiile raportate în literatură.



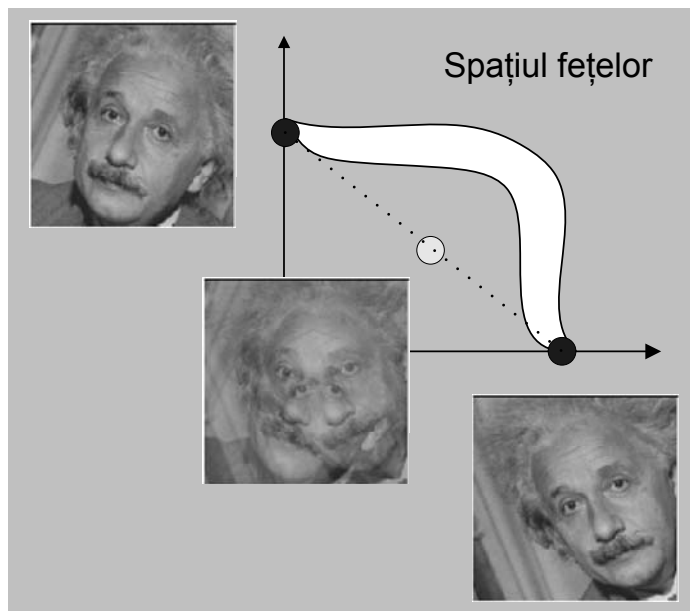
Fig. 1.2 Aspecte specifice aplicațiilor de recunoaștere/verificare a fețelor

### *Ce reprezintă o față?*

Dacă transformăm o imagine de dimensiuni  $N \times N$  pixeli într-un vector de dimensiune  $N^2$ , acesta poate fi privit ca un punct într-un spațiu  $N^2$  - dimensional. Imaginile reprezentând fețe umane ocupă doar un mic subspațiu al acestui spațiu multidimensional, cu caracteristici specifice. Se poate arăta că modificările

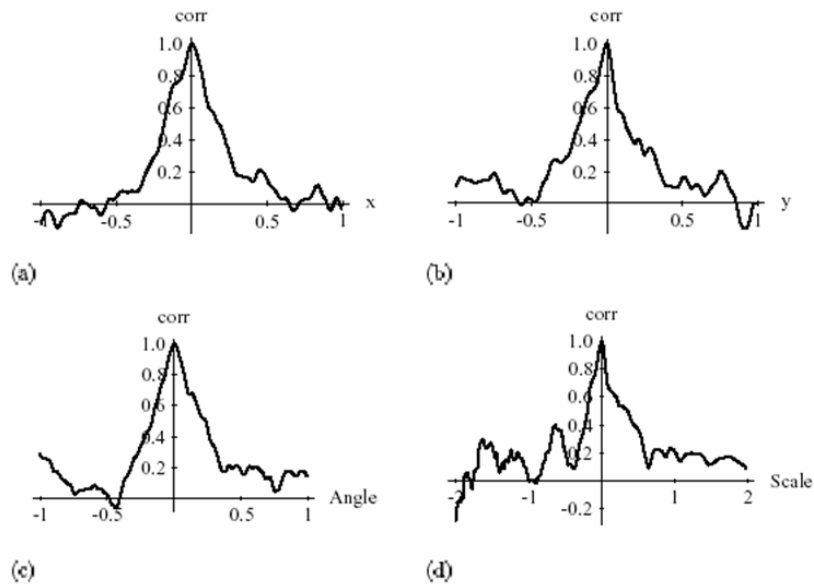
obișnuite precum translația, rotația, sau schimbarea nivelului de iluminare, atunci când au amplitudine mică, definesc simple subregiuni compacte din “subspațiul fețelor”. Pentru transformări mai generale – rotații mari, acoperire parțială, schimbări de scală – subregiunile feței devin non-convexe. Un exemplu în acest sens este dat în Fig. 1.3, arătând că o combinație liniară dintre o față și versiunea sa rotită nu este o imagine validă a unei fețe. În plus, asemănarea dintre imaginea originală și versiunea sa transformată se degradează rapid odată cu creșterea amplitudinii parametrilor ce descriu transformările respective. În Fig. 1.4 este ilustrată această observație, unde funcția de corelație este aleasă ca o măsură a asemănării dintre vectorii comparați [22].

Datorită atributului “non-convex”, metodele de procesare a feței bazate pe o singură imagine sunt limitate în performanță. Apare acum o întrebare interesantă: de câte imagini bidimensionale (2-D) este nevoie pentru a reprezenta o față, indiferent de orientarea acesteia? Câteva răspunsuri au fost date în literatură, indicând faptul că 4 sau 5 imagini sunt suficiente pentru a acoperi toate unghiurile de vedere (alegerea lor este dependentă de un set de constrângeri) [3].



**Fig. 1.3** Exemplu indicând non-convexitatea spațiului fețelor: imaginea originală, versiunea rotită a aceleiași fețe, combinație liniară între fețe





**Fig. 1.4** Asemănarea dintre imaginea originală a feței și versiunea ei transformată se deteriorează rapid pe măsura creșterii valorii parametrilor ce definesc transformarea: a) translație pe axa X; b) translație pe axa Y; c) rotație; d) modificarea scalei

### *Care este cea mai distinctivă informație?*

După cum s-a menționat anterior, există 2 strategii majore implicate în obținerea "semnăturii" feței, anume:

- a) tehnica geometrică, bazată pe extracția pozițiilor relative și a altor parametri dependenți de puncte particulare precum ochii, colțurile gurii, nasul și bărbia
- b) tehnica bazată pe modele (*templates*), în care matricele reprezentând valorile de intensitate luminoasă a pixelilor care alcătuiesc fața de test și, respectiv, cea de referință sunt procesate convenabil, urmând ca rezultatele să fie comparate utilizând un anumit tip de măsură a similitudinii (câteodată sunt folosite mai multe modele pentru fiecare față).

Alegerea procedurii specific de extragere a "semnăturii" este dictată de capacitatea distinctivă a informației rezultate. Din moment ce imaginile feței sunt reprezentate în mod obișnuit printr-o matrice de dimensiuni considerabile, se folosesc diverse metode de compresie, liniare sau neliniare, precum Analiza pe Componente Principale (*Principal Components Analysis – PCA*) pentru obținerea

unei reprezentări de dimensiuni mult reduse a imaginii originale, fără o pierdere semnificativă a calității. Mai mult, Analiza Discriminatorie Liniară (*Linear Discriminant Analysis* – LDA) este adesea aleasă pentru a identifica direcțiile din “spațiul fețelor” de-a lungul cărora separarea „semnăturilor” este maximă.

O abordare originală constă în organizarea spațiului fețelor în mod ierarhic, partiționarea urmând o structură de tip arbore. Merită notat că tipul de codare a imaginilor originale poate fi de asemenea influențat de necesitatea de a asigura invarianța la transformări comune. O direcție recentă de cercetare este legată de așa-numitul “*optical flow*” [15], bazat pe reprezentarea variației (gradientului) în intensitate dintre două imagini distincte ale aceleiași persoane.

O observație foarte interesantă este legată de reprezentarea imaginilor în domeniul frecvență: informația de fază este mult mai importantă pentru discriminare (separare) decât cea de amplitudine, după cum sugerează exemplul din Fig. 1.5 [21].



**Fig. 1.5** a) imagini originale; b) aceeași amplitudine și fază aleatoare; imagini reconstruite utilizând informația de amplitudine corectă și faza corespunzătoare celeilalte persoane

*Procesare locală sau globală ?*

Modularitatea reprezintă o caracteristică fundamentală a creierului uman, care a inspirat numeroase studii ale cercetătorilor din domeniul neurofiziologiei, dar și al rețelelor neurale artificiale. Dintre avantajele față de structurile nemodulare amintim timpul redus de antrenare, degradarea mai lentă a performanțelor odată cu creșterea complexității aplicației considerate, înțelegerea mai comodă a rolului jucat de diversele componente ale unui sistem complex. În cazul aplicațiilor de recunoaștere a fețelor, abordarea modulară este strâns legată de tehnicile care își propun extragerea unei semnături bazate pe *procesare locală*, efectuată în jurul unor puncte reprezentative (ochi, nas, gură). În această categorie sunt incluse metodele denumite *Elastic Bunch Graph Matching* [25] și *Local Feature Analysis* [19], precum și cele bazate pe utilizarea filtrelor Gabor [16]. La polul opus se situează așa-numitele *metode holistice*, care extrag semnături luând în considerare întreaga față. În plus, abordarea modulară a fost utilizată și pentru a compensa problemele cauzate de orientarea variabilă a capului în raport cu aparatul de fotografiat [20].

*Cum se asigură invarianța în raport cu transformări elementare ?*

Aplicațiile de recunoaștere a feței se confruntă cu multe surse de variabilitate, printre care putem enumera orientarea în raport cu aparatul de fotografiat, nivelul de iluminare, acoperirea parțială, expresia feței, intervalul de timp care separă momentele de achiziție a unor imagini distincte ale unei aceleiași persoane.

Variațiile datorate nivelului de iluminare au atras o atenție deosebită, datorită necesității de a beneficia de metode robuste de recunoaștere inclusiv în cazul imaginilor captate în mediul exterior, nu numai al celor obținute în încăperi cu condiții controlate. Astfel, au fost elaborate analize teoretice asupra subspațiilor generate de imagini supuse iluminării variabile, identificându-se numărul minim de surse de lumină distincte necesare pentru a modela o față indiferent de direcția de iluminare [3].

În plus, sistemele automate de procesare a fețelor ar trebui să tolereze între anumite limite și prezența unui set de transformări geometrice elementare precum

translațiile, rotațiile în planul imaginii sau schimbările de scală. Au fost propuse două abordări majore pentru atingerea acestui obiectiv:

a) algoritmi de procesare care asigură extragerea din imaginile originale a unor *trăsături invariante* la astfel de transformări. Drept exemple din această categorie putem enumera:

- funcții de autocorelație de ordin superior (invariantă la translație) [10] definite prin relația următoare, unde  $I(\mathbf{r})$  desemnează imaginea, iar  $\mathbf{a}_i$  vectori de deplasare în planul P al imaginii:

$$f(\mathbf{a}_1, \dots, \mathbf{a}_N) = \int_P I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1) \dots I(\mathbf{r} + \mathbf{a}_N) d\mathbf{r} \quad (1.1)$$

- modulul transformatei Fourier (bidimensionale) a imaginii originale (invariantă la translație)

- reprezentarea în coordonate polare a imaginii originale (în realitate, distanța unui punct față de originea sistemului de coordonate se reprezintă pe o scală logaritmică, astfel încât reprezentarea este de tip log-polar [11]). Ca urmare, rotația în planul imaginii, respectiv modificarea scalei de vizualizare (*zoom-ul*), se vor reduce la apariția unor translații de-a lungul uneia dintre cele 2 axe. Pentru a asigura invarianța în raport cu aceste două transformări va fi necesar să extragem o informație invariantă doar în raport cu translația, de exemplu folosind una dintre metodele enumerate anterior. Metoda denumită *spectroface* [12] extrage trăsături invariante în raport cu toate cele 3 tipuri de transformări elementare, combinând reprezentarea log-polară cu utilizarea transformatei Fourier.

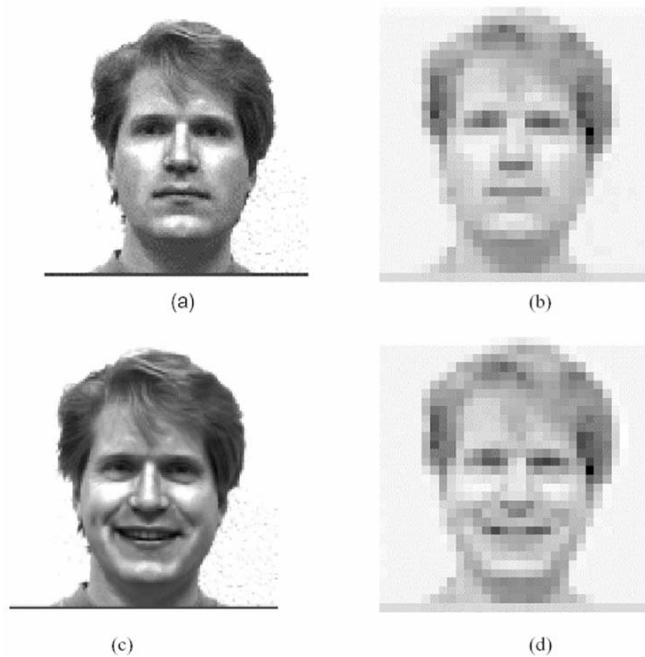
b) efectul aplicării unei transformări elementare poate fi folosit explicit în definiția distanței utilizate pentru aprecierea similitudinii dintre imagini. Această abordare este ilustrată prin metodele bazate pe așa-numita *distanță de tip tangentă* [23] și, respectiv, modelele de deformare (*Active Appearance Models*) [28].

*Cum depind performanțele de componența spectrală și rezoluția de reprezentare ?*

O serie de studii indică faptul că frecvențele spațiale joase, respectiv cele înalte joacă roluri diferite în privința recunoașterii și concluzionează că:

- componentele de joasă frecvență sunt suficiente pentru recunoaștere
- efectul expresiei faciale poate fi atenuat prin eliminarea componentelor de frecvență înaltă

De obicei tehnicile bazate pe Transformata Wavelet Discretă (DWT) sunt folosite pentru a extrage semnături de frecvență joasă a feței, ca în exemplul din Fig. 1.6 [11, 12]. Influența rezoluției de reprezentare asupra ratei de recunoaștere a fost de asemenea analizată în literatură și este interesant de semnalat faptul că o rezoluție de numai 32x32 pixeli este suficientă pentru a obține performanțe rezonabile (în fapt, anumite aplicații precum simpla identificare a genului bărbat/femeie necesită chiar rezoluții mai mici).



**Fig. 1.6** a) expresie normală; b) reprezentare de joasă frecvență a expresiei normale după aplicarea DWT; c) expresie fericită; d) reprezentare de joasă frecvență a expresiei fericite după aplicarea DWT

### *Ce clasificator să utilizăm ?*

Una dintre cele mai surprinzătoare observații indicate în literatură arată că distanțele dintre imaginile unei aceleași persoane care diferă prin condițiile de iluminare sau prin orientarea în raport cu aparatul de fotografiat sunt **mai mari**

decât distanțele dintre imagini ale unor persoane diferite, prelevate însă în aceleași condiții. Aceste observații transformă recunoșterea feței într-o sarcină foarte dificilă, iar proiectarea clasificatorului nu este deloc simplă. Sunt utilizate cu precădere următoarele două strategii:

a) regula “celui mai apropiat vecin” (*nearest-neighbor rule*): când o imagine test este prezentată la intrarea sistemului de recunoaștere a feței (care a fost antrenat anterior cu un set de imagini prototip), ieșirea este definită prin eticheta prototipului care este cel mai apropiat (conform unei distanțe (metrici) adecvate) de imaginea de test. În mod obișnuit comparația este făcută luându-se în considerare nu imaginile originale, ci “semnăturile” extrase prin proceduri specifice, iar alegerea concretă a distanței poate fi adaptată aplicației considerate. Uneori, o strategie ierarhică se poate dovedi superioară din punctul de vedere al ratei de recunoaștere sau robusteții clasificării. De asemenea, se poate folosi un ansamblu de clasificatoare distincte având tipuri diferite de date de intrare.

b) rețele neurale recurente [4]: sistemele dinamice neliniare pot avea, în anumite condiții, *puncte de echilibru stabil* în poziții predefinite. Acelor puncte de echilibru le vor corespunde un set de imagini prototip din baza de date de antrenare. Imaginile unor persoane de test, văzute ca versiuni zgomotoase, incomplete sau distorsionate ale imaginilor prototip, vor servi drept *condiții inițiale* aparținând *bazinelor de atracție* ale acestor puncte de echilibru, iar dinamica sistemului va conduce în final la stabilizarea ieșirii chiar la valorile vectorilor prototip. Această abordare a fost folosită cu precădere în legătură cu strategia de *asociere temporală* a imaginilor (vederi ușor diferite ale aceleiași persoane sunt de regulă obținute în strictă succesiune temporală). Avantajul unei asemenea abordări în comparație cu regula “celui mai apropiat vecin” constă în faptul că nu mai este necesar calculul unei distanțe în raport cu toate imaginile prototip (aspect dificil dacă baza de date prototip are dimensiune mare), însă dificultățile sunt legate de necesitatea de stocare doar a punctelor de echilibru dorite, nu și a unora false, precum și de capacitatea limitată de memorare a imaginilor prototip.

O abordare interesantă este legată de așa-numitele *rețele neurale autoasociative*, a căror arhitectură este prezentată în Fig. 1.7: o rețea multistrat având același

număr de intrări și ieșiri, precum și *aceleași date* livrate ca intrări și ieșiri dorite, este antrenată folosind exemple aparținând *unei singure clase* (diferite imagini ale unei aceleași persoane). În faza de testare, o imagine nouă este aplicată la intrare, iar *eroarea de reconstrucție* (distanța euclidiană dintre intrare și ieșire) este folosită ca informație discriminativă: dacă imaginea aparține clasei corecte (imaginea de test corespunde persoanei ale cărei fotografii au fost utilizate pentru determinarea valorilor parametrilor rețelei neurale) această eroare va fi mult mai mică decât în cazul altor clase. Această abordare este folositoare în mod particular în aplicațiile de verificare și rezolvă problemele datorate existenței unui număr limitat de fotografii corespunzătoare unei aceleași persoane.

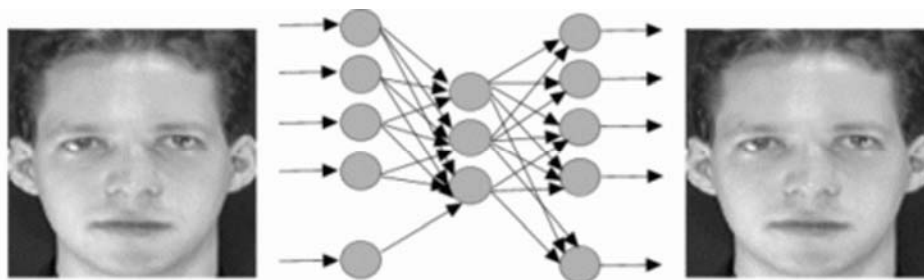


Fig. 1.7 Arhitectura unei rețele neurale autoasociative

## 1.2 Algoritmi de recunoaștere a fețelor

În literatură au fost propuse numeroase tehnici de recunoaștere/autentificare a fețelor, cu grade de complexitate, constrângeri, performanțe și arii de aplicabilitate foarte diverse. În mod special, identificarea, extragerea și ierarhizarea setului de trăsături semnificative care va constitui „semnătura” fiecărei fețe supuse analizei continuă să reprezinte un subiect de larg interes. Paleta foarte largă de opțiuni din această categorie poate fi clasificată ținând cont de o serie de criterii precum natura informațiilor extrase (distingem trăsături rezultate din geometria specifică fețelor, respectiv în urma aplicării unei proceduri de proiecție pe subspații liniare sau neliniare specifice), aria imaginii supuse procesării (vom regăsi metode, denumite *holistice*, care prelucrează întreaga imagine, respectiv soluții orientate pe zone limitate situate în jurul punctelor fiduciale), gradul de localizare spațială a

trăsăturilor, natura reală (eventual binară) sau complexă a semnăturii, modularitatea arhitecturii, robustețea în raport cu numeroasele surse de variabilitate specifice acestor aplicații. De multe ori alegerea metodei de extragere a trăsăturilor semnificative este strâns legată de tipul clasificatorului ce urmează a fi utilizat în vederea furnizării deciziei. Mai mult, dimensiunea și particularitățile bazei de date disponibile impun constrângeri suplimentare în privința alegerii soluției adecvate.

În cele ce urmează vom trece în revistă câteva dintre tehnicile semnificative descrise în literatură, menționând aspectele teoretice, avantajele și dezavantajele, precum și codul MATLAB corespunzător.

#### *Analiza pe Componente Principale (PCA)*

De regulă, bazele de date folosite în experimentele de recunoaștere a fețelor conțin imagini de dimensiune foarte mare. O astfel de “risipă” de resurse, valabilă de altfel și în cazul semnalelor vocale sau al altor imagini naturale, conduce la o robustețe semnificativă, care permite receptarea corectă a informației transmise, chiar în condițiile în care aceasta este afectată de zgomot, este distorsionată sau incompletă. Pe de altă parte, dimensiunile mari complică semnificativ implementarea practică a diverselor tehnici de procesare, cresc volumul de calcul și, în plus, necesită existența unui număr sporit de imagini în baza de date cu care se operează (dacă imaginile originale sunt văzute ca puncte într-un spațiu multidimensional, cu cât dimensiunea spațiului este mai mare, cu atât mai multe puncte sunt necesare pentru a asigura o “acoperire” mai bună a întregului spațiu, în vederea asigurării unei aproximări adecvate a densității reale de repartiție a tuturor punctelor reprezentând imagini valide de fețe umane). În acest context, se dovedesc utile *tehnicele de compresie*, folosite pentru a reduce dimensiunea datelor originale, în condițiile unor pierderi de informație (inevitabile) cât mai mici.

Tehnicile de compresie uzuale aparțin de regulă uneia din următoarele 3 categorii: a) codare (liniar) predictivă; b) calcul de transformate liniare; c) cuantizare vectorială. În cele ce urmează vom introduce una dintre cele mai cunoscute metode de compresie aparținând celei de a doua clase, denumită Analiza pe Componente Principale (*Principal Component Analysis – PCA*) sau transformata Karhunen-Loeve. Astfel, să considerăm că avem la dispoziție  $N$  vectori  $\mathbf{x}_n$  aparținând unui



spațiu vectorial de dimensiune  $D$  cărora li se asociază, prin intermediul unei transformate liniare descrise de matricea  $[\mathbf{W}]_{M \times N}$ , un set de vectori de dimensiune mai mică  $M < D$ , conform relației:

$$\mathbf{y}_n = \mathbf{W}\mathbf{x}_n, \quad n = 1 \dots N \quad (1.2)$$

În spațiul original, vectorul  $\mathbf{x}_n$  poate fi scris sub forma unei combinații liniare a elementelor unei baze ortonormate  $\{\mathbf{u}_i, i = 1 \dots D\}$  sub forma:

$$\mathbf{x}_n = \sum_{i=1}^D w_{i,n} \mathbf{u}_i, \quad n = 1 \dots N \quad (1.3)$$

în care coeficienții  $w_{i,n}$  se calculează cu relația:

$$w_{i,n} = \mathbf{u}_i^T \mathbf{x}_n, \quad n = 1 \dots N \quad (1.4)$$

Să presupunem că dorim să reținem din descompunerea (1.3) numai un set redus de  $M$  termeni, iar restul coeficienților  $w_{i,n}$  să fie înlocuiți *cu valori constante*:

$$\tilde{\mathbf{x}}_n = \sum_{i=1}^M w_{i,n} \mathbf{u}_i + \sum_{i=M+1}^D c_i \mathbf{u}_i, \quad n = 1 \dots N \quad (1.5)$$

Ideea fundamentală a metodei PCA este de a găsi baza particulară care asigură **minimizarea erorii pătratice medii** dintre seturile de vectori  $\mathbf{x}_n$  și  $\tilde{\mathbf{x}}_n$ :

$$E = \frac{1}{2} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2 = \frac{1}{2} \sum_{n=1}^N \sum_{i=M+1}^D (w_{i,n} - c_i)^2 \quad (1.6)$$

Un rezultat fundamental cunoscut sub denumirea de **transformata Karhunen-Loeve** (KLT) demonstrează că această bază este formată din *vectorii proprii ai matricii de autocovarianță* a vectorilor  $\mathbf{x}_n$ :

$$S = E\{(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T\} \approx \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T \quad (1.7)$$

unde  $\bar{\mathbf{x}}$  desemnează valoarea estimată (pe baza setului finit de  $N$  exemple) a valorii medii a procesului aleator reprezentat de vectorii  $\mathbf{x}_n$ . În plus, valorile constantelor  $c_i$  sunt date de relația:

$$c_i = \frac{1}{N} \sum_{n=1}^N w_{i,n} = \mathbf{u}_i^T \bar{\mathbf{x}} \quad (1.8)$$

Utilizarea acestor rezultate pentru compresia de date presupune parcurgerea succesivă a câtorva pași: a) calculul valorii medii  $\bar{\mathbf{x}}$  a vectorilor supuși analizei; b) determinarea valorilor  $\lambda_i$  și vectorilor proprii  $\mathbf{u}_i$  ai matricii de autocovarianță  $S$ ; c) aplicarea relației (1.5) pentru un număr  $M < D$  de vectori proprii corespunzători celor mai mari valori proprii ale matricii de autocovarianță  $S$  (așa-numitele componente principale). În multe aplicații practice se dovedește că cea mai mare parte a energiei semnalelor originale se regăsește acumulată numai într-un număr redus de componente principale, iar modalitatea particulară de selecție a acestora este justificată (conform unui calcul elementar) de expresia erorii pătratice medii:

$$E = \frac{1}{2} \sum_{i=M+1}^D \lambda_i \quad (1.9)$$

Din punct de vedere practic apar dificultăți datorate complexității calculului și memoriei necesare estimării matricii de autocovarianță  $S$  și a valorilor/vectorilor proprii ale acesteia. Una dintre soluțiile adoptate este de a înlocui baza ortonormată corespunzătoare transformatei KLT cu cea definită de Transformata Cosinus Discretă (DCT)<sup>1</sup>. Pe de altă parte au fost prezentate în literatură o serie de rețele neurale, cele mai multe liniare, capabile să implementeze metoda PCA într-o manieră recursivă, eliminând astfel și dezavantajul variantei standard de a necesita reluarea întregii proceduri de calcul la fiecare apariție a unui vector nou în baza de date.

Metoda PCA descrisă anterior a fost aplicată cu succes la recunoașterea fețelor, sub denumirea generică *Eigenfaces* [24]. ”Materia primă” o constituie ansamblul imaginilor disponibile în baza de date, formate din matrici cu valori reale (eventual – binare). Fiecare astfel de matrice, presupusă de dimensiune  $(M \times N)$ , este mai întâi transformată într-un vector de aceeași lungime, prin concatenarea coloanelor corespunzătoare. Algoritmul de procesare presupune parcurgerea următorilor pași:

---

<sup>1</sup> Această metodă este implementată în standardul JPEG (*Joint Photographic Expert Group*)

a) Se calculează valoarea medie a imaginilor care formează setul de antrenare (presupus a avea K fotografii):

$$\bar{I} = \frac{\sum_{j=1}^K I_j}{K} \quad (1.10)$$

și se „centrează” imaginile originale (se aduc la valoare medie nulă):

$$I_j^{centrat} = I_j - \bar{I} \quad (1.11)$$

b) Se calculează așa-numita *scatter matrix*, care reprezintă aproximarea matricii de covarianță a imaginilor din baza de date (aproximarea este cu atât mai bună cu cât avem mai multe imagini la dispoziție):

$$S = \frac{1}{K} A \cdot A^T \quad (1.12)$$

unde matricea A are pe coloane câte o fotografie centrată:

$$A = \begin{bmatrix} I_1^{centrat} & I_2^{centrat} & \dots & I_k^{centrat} \end{bmatrix}_{(M \times N) \times k} \quad (1.13)$$

Matricea S este simetrică și are dimensiuni  $(M \cdot N) \times (M \cdot N)$ .

c) Se calculează valorile și vectorii proprii ai matricii S (vectorii proprii ai matricii S în cazul lucrului cu imagini reprezentând fețe poartă denumirea *Eigenfaces*).

**Observații:**

a) se poate utiliza un artificiu care reduce volumul de calcul: se calculează valorile și vectorii proprii ai matricii  $A^T A$  și apoi se folosește relația dintre aceștia din urmă și vectorii proprii ai matricii S.

b) valorile proprii ale matricii S sunt întotdeauna pozitive deoarece S este reală și simetrică.

d) Se ordonează valorile proprii ale matricii S în sens descrescător. Se trasează un grafic care exprimă pierderea de informație în raport cu factorul de compresie.

Astfel, dacă notăm cu  $\lambda_i$ ,  $i = 1 \dots \overline{(M \times N)}$ , valorile proprii sortate ale matricii  $S$ , graficul anterior se referă la raportul (pe abscisă avem  $j = 1 \dots \overline{(M \times N)}$ ):

$$\frac{\sum_{i=1}^j \lambda_i}{\sum_{j=1}^{\overline{(M \times N)}} \lambda_j} \quad (1.14)$$

Graficul anterior permite estimarea numărului de valori și vectori proprii considerați semnificativi (adică aceia care păstrează cea mai mare parte din energia imaginilor originale).

e) Se proiectează imaginile (centrate) originale pe spațiul descris de vectorii proprii reprezentativi (tipic aceștia sunt în număr de 5-10% din numărul total). Proiecția constă de fapt în efectuarea produsului scalar dintre fiecare imagine originală și o matrice având drept coloane numai vectorii proprii semnificativi.

Pentru fiecare imagine (centrată)  $I_j^{centrat}$  se obține proiecția pe baza relației:

$$W_j^T = I_j^{centratT} \cdot V_{PCA} \quad (1.15)$$

$$V_{PCA} = [E_1 E_2 E_3 \dots E_{N_{\max}}] \quad (1.16)$$

unde  $N_{\max}$  este numărul maxim de vectori proprii reținuți,  $E_j$  sunt vectorii proprii semnificativi, iar vectorii  $W_j$  au dimensiunea  $(N_{\max} \times 1)$  și pot fi priviți ca „semnăturile” asociate imaginilor originale.

f) Clasificarea imaginilor test presupune mai întâi determinarea „semnăturii” fiecărei imagini în raport cu subspațiul determinat anterior (și care depinde exclusiv de imaginile din setul de antrenare!) și găsirea acelei imagini din baza de date de antrenare a cărei semnătură este cea mai apropiată de semnătura imaginii de test. Aprecierea similitudinii dintre astfel de perechi de imagini se realizează folosind o metrică convenabil aleasă. Opțiunea uzuală este distanța Euclideană (L2), însă se

pot utiliza și alte măsuri precum funcția de autocorelație, cosinusul unghiului dintre 2 vectori sau distanța Mahalanobis. Definițiile acestora se prezintă mai jos:

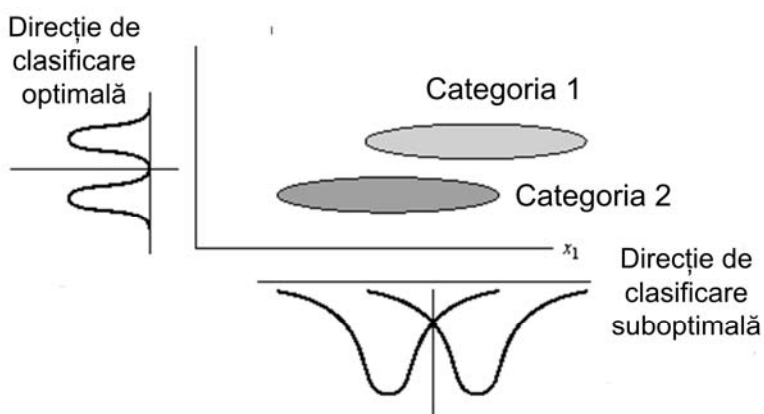
$$\begin{aligned}
 \text{Distanța Euclideană:} \quad d_{L_2}(x, y) &= \sum_{i=1}^D (x_i - y_i)^2 \\
 \text{Distanța Manhattan:} \quad d_{L_1}(x, y) &= \sum_{i=1}^D |x_i - y_i| \\
 \text{Funcția de intercorelație:} \quad C[k] &= \sum_{k=-D}^D x_i \cdot y_{i+k} \quad (1.17) \\
 \text{Cosinusul unghiului dintre vectori:} \quad \cos(x, y) &= \frac{x \cdot y}{\|x\| \cdot \|y\|} \\
 \text{Distanța Mahalanobis:} \quad d_M(x, y) &= \frac{\sum_{i=1}^D x_i \cdot y_i}{\sqrt{\lambda_i}}
 \end{aligned}$$

Principalul avantaj al metodei PCA constă în simplitatea sa. Există posibilitatea de a calcula vectorii proprii principali nu numai în variantă *off-line* prin procedura algebrică descrisă anterior ci și *on-line*, folosind anumite rețele neurale artificiale pentru a ajusta în mod iterativ valorile acestora, pe măsură ce se aplică date noi la intrare. În cazul aplicării metodei PCA la recunoașterea fețelor, trebuie menționate și o serie de dezavantaje, care limitează performanțele sistemelor bazate pe acest instrument de analiză:

- principalul dezavantaj constă în faptul că ignorarea componentelor care contribuie puțin la energia imaginilor originale nu conduce automat și la îmbunătățirea separării (discriminării) dintre diversele subclase de imagini aparținând unor persoane diferite (altfel spus, minimizarea erorii de reprezentare nu înseamnă neapărat îmbunătățirea performanțelor de clasificare!). Un exemplu intuitiv este prezentat în Fig. 1.8, care indică faptul că cele două clase pot fi separate comod efectuând proiecția pe componenta principală *mai puțin semnificativă*.
- un alt dezavantaj al tehnicii PCA îl reprezintă caracterul global al acesteia, în sensul că matricea de covarianță (deci și valorile și vectorii proprii) este dedusă

luând în considerare toate imaginile disponibile, care pot să conțină o foarte mare variabilitate din punct de vedere al nivelului de iluminare, orientării, fundalului. Acest dezavantaj poate fi compensat parțial utilizând arhitecturi modulare, în care tehnica PCA se aplică pe subseturi de imagini care au caracteristici comune (de exemplu, se calculează seturi de vectori proprii distincți pentru clasa imaginilor reprezentând vederi frontale, respectiv vederi înclinate la  $\pm 30^\circ$ ,  $\pm 40^\circ$ ). De asemenea, se pot utiliza arhitecturi în care fiecare modul este “specializat” să recunoască numai imaginile reprezentând o aceeași persoană.

- în cazul particular al analizei imaginilor reprezentând fețe, metoda PCA s-a dovedit extrem de sensibilă în raport cu normalizarea poziției acestora: translații pe orizontală/verticală cu numai câțiva pixeli, mici rotații sau variații de scală pot altera semnificativ „semnăturile” extrase și, în consecință, pot degrada performanțele de recunoaștere/autentificare. Ca urmare, este necesară o etapă (manuală sau automată) de normalizare a aspectului imaginilor, care să preceadă aplicarea propriu-zisă a PCA. Alternativ, au fost propuse variante ale metodei standard, care încearcă să compenseze acest dezavantaj prin filtrarea adecvată a imaginilor originale. În particular, uneori se preferă ignorarea proiecției de-a lungul primului vector propriu semnificativ, deoarece de regulă acesta este sensibil la nivelul global de iluminare al imaginilor.



**Fig. 1.8** Metoda PCA nu asigură întotdeauna maximizarea discriminabilității: direcția de clasificare optimală corespunde vectorului propriu mai puțin semnificativ

**pca.m:** Funcție MATLAB care implementează algoritmul PCA

```

function [eig_vectors, eig_values] = pca(A, numvecs);
% Functia returneaza un numar de vectori proprii semnificativi ai matricii de covarianta
% asociate matricii A, folosind algoritmul lui Turk & Pentland [24] pentru cazul tipic in care
% numarul de linii ale lui A este mult mai mare decit cel al coloanelor

% A - matricea datelor de intrare (fiecare coloana reprezinta un vector distinct)
% numvecs - numarul vectorilor proprii selectati
% eig_vectors - matricea vectorilor proprii (organizati pe coloane)
% eig_values - valorile proprii

nexamp = size(A,2);
mean_A = mean(A)'; % valoarea medie

% Se centreaza vectorii care formeaza matricea A
for i = 1:nexamp
    A(:,i) = A(:,i) - mean_A;
end

CovMat=A*A'; % matricea de covarianta asociata lui A
L = A'*A; % matricea folosita de catre Turk & Pentland
[Vectors,Values] = eig(L);

% Sortarea vectorilor proprii in functie de valorile proprii
[Vectors,Values] = sortem(Vectors,Values);

% Turk & Pentland: vectorii proprii ai lui L se transforma in cei ai lui CovMat
Vectors = A*Vectors;

% Extragerea valorilor proprii
Values = diag(Values);
Values = Values / (nexamp-1);
% Normalizarea vectorilor proprii, eliminarea valorilor proprii nesemnificative
num_good = 0;
for i = 1:nexamp
    Vectors(:,i) = Vectors(:,i)/norm(Vectors(:,i));
    if Values(i) < 0.00001
        Values(i) = 0;
        Vectors(:,i) = zeros(size(Vectors,1),1);
    else
        num_good = num_good + 1;
    end
end
if (numvecs > num_good)
    sprintf(1,'Warning: numvecs is %d; only %d exist.\n',numvecs,num_good);
    numvecs = num_good;
end
Vectors = Vectors(:,1:numvecs);

```

```

%=====
function [NV,ND] = sortem(V,D);
%=====
% Functia sorteaza coloanele matricii V in functie de valorile de pe diagonala principala a
% matricii D
dvec = diag(D);
NV = zeros(size(V));
[dvec,index_dv] = sort(dvec);
index_dv = flipud(index_dv);
for i = 1:size(D,1)
    ND(i,i) = D(index_dv(i),index_dv(i));
    NV(:,i) = V(:,index_dv(i));
end

```

### *Analiza Discriminatorie Liniară (LDA)*

O observație importantă menționată anterior sublinia faptul că proiecția datelor de lucru pe subspațiul generat de setul de vectori proprii semnificativi ai matricii de autocorelație, deși minimizează eroarea de reconstrucție, nu oferă automat și optimizarea performanțelor de clasificare. Pentru a identifica direcțiile din spațiu de-a lungul cărora ar fi indicat să efectuăm proiecția pentru a maximiza separarea (discriminabilitatea) datelor procesate se poate utiliza o altă tehnică liniară de natură statistică, cunoscută sub denumirea *Linear Discriminant Analysis (LDA)*. Pentru a ilustra această metodă, să considerăm cazul mai simplu al clasificării unui set de vectori  $\mathbf{x}$  în 2 categorii distincte, având fiecare  $N_1$ , respectiv  $N_2$  exemplare. Să considerăm o direcție din spațiu, definită de un vector  $\mathbf{w}$  cu aceeași dimensiune ca și vectorii  $\mathbf{x}$ , de-a lungul căreia să proiectăm datele de lucru pe baza relației  $y = \mathbf{w}^T \cdot \mathbf{x}$ , decizia privind apartenența vectorului  $\mathbf{x}$  la una dintre cele 2 categorii urmând a fi luată prin compararea mărimii scalare  $y$  cu o valoare de prag convenabil aleasă. Ne propunem să identificăm acea direcție particulară care asigură maximizarea ratei de clasificare corectă, altfel spus direcția de-a lungul căreia separarea dintre clase este cea mai mare. Una dintre cele mai naturale idei este de a considera direcția care mărește separabilitatea proiecțiilor valorilor medii ale celor 2 clase, cu alte cuvinte ar trebui să maximizăm expresia:

$$m_2 - m_1 = \mathbf{w}^T \cdot (\mathbf{m}_2 - \mathbf{m}_1) \quad (1.18)$$

unde  $\mathbf{m}_1$  și  $\mathbf{m}_2$  desemnează valorile medii:



$$\mathbf{m}_1 = \frac{1}{N_1} \sum_{n \in C_1} \mathbf{x}_n ; \quad \mathbf{m}_2 = \frac{1}{N_2} \sum_{n \in C_2} \mathbf{x}_n \quad (1.19)$$

Deoarece expresia (1.18) poate fi făcută oricât de mare ‘‘jonglând’’ din amplitudinea vectorului  $\mathbf{w}$ , este necesar să introducem o constrângere suplimentară, de exemplu să impunem ca norma acestuia să fie constantă. Utilizând binecunoscuta metodă a multiplicatorilor lui Lagrange, se poate arăta că soluția noii probleme de optimizare (cu constrângeri) conduce la  $\mathbf{w} \sim (\mathbf{m}_2 - \mathbf{m}_1)$ .

Totuși, este posibil ca direcția de maximă discriminabilitate să nu fie neapărat cea care asigură separarea optimă a proiecțiilor valorilor medii, în special în cazurile în care una sau ambele clase conțin date foarte ‘‘împrăștiate’’ în spațiu. Ca urmare, o soluție mai bună este cea care își propune maximizarea unei funcții definite ca raportul dintre proiecțiile valorilor medii și dispersiile datelor de-a lungul direcției vectorului  $\mathbf{w}$ :

$$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{\sigma_1^2 + \sigma_2^2} = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (1.20)$$

unde matricile  $\mathbf{S}_B$  și  $\mathbf{S}_W$  sunt definite prin:

$$\begin{aligned} \mathbf{S}_B &= (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \\ \mathbf{S}_W &= \sum_{n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^T + \sum_{n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^T \end{aligned} \quad (1.21)$$

Se poate arăta că maximizarea funcției  $J(\mathbf{w})$  conduce la concluzia:

$$\mathbf{w} \sim \mathbf{S}_W^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \quad (1.22)$$

În cazul particular în care matricea  $\mathbf{S}_W$  este proporțională cu o matrice unitate (se spune că matricea este *izotropă*, adică împrăștierea datelor este uniformă în toate direcțiile), condiția anterioară se reduce din nou la  $\mathbf{w} \sim (\mathbf{m}_2 - \mathbf{m}_1)$ .

Metoda LDA poate fi generalizată pentru probleme de clasificare cu un număr oarecare de categorii  $C \geq 2$ , cu observația că numărul de direcții discriminatorii posibil de determinat este egal cu  $(C-1)$ . Această constrângere poate reprezenta un dezavantaj major pentru extragerea unor ‘‘semnături’’ semnificative în condițiile în

care avem la dispoziție date de dimensiune mare aparținând unui număr limitat de clase, motiv pentru care de multe ori aplicarea metodei LDA este precedată de o etapă prealabilă de reducere a dimensionalității vectorilor originali, de obicei bazată pe aplicarea tehnicii PCA. Mai mult, dacă numărul de vectori disponibili este redus (iar acest lucru este posibil în aplicații biometrice unde, de exemplu, putem avea numai una sau două fotografii ale unei aceleiași persoane!), estimarea valorilor medii și, în continuare, a matricilor  $S_B$  și  $S_W$  este nerelevantă din punct de vedere statistic. Un alt dezavantaj, pe care îl prezintă ambele tehnici descrise anterior, se referă la faptul că identificarea setului de trăsături semnificative se face folosind exclusiv datele de intrare, nu și eticheta corespunzătoare categoriei cărora acestea aparțin. În final, să menționăm că în literatură au fost propuse și utilizate cu succes unele variante neliniare ale acestor metode, de exemplu bazate pe arhitecturi speciale de rețele neurale artificiale.

**lda.m:** Funcție MATLAB care implementează algoritmul LDA

```
function [fisher_basis]=lda(data, NumClasses, ImgsPerClass);
% data - matricea de date (organizate pe coloane)
% NumClasses - numarul de categorii
% ImgsPerClass - numarul de vectori din fiecare clasa (presupus identic pentru simplitate)
% fisher_basis - matrice ale carei coloane reprezinta directiile de maxima discriminabilitate

N=size(data_train,1); % dimensiunea vectorilor de intrare
% Se calculeaza vectorul mediu al fiecarei clase
mn = mean(data)';
for i = 1:NumClasses
    m(:,i) = mean(data(:,(i-1)*ImgsPerClass+1:(i-1)*ImgsPerClass+ImgsPerClass))';
    msm(:,i) = m(:,i) - mn;
end
% Se centreaza toate imaginile
for i=1:NumClasses*ImgsPerClass
    msc(:,i) = data(:,i) - m(:,double(floor((i-1)/ImgsPerClass))+1);
end

sw = zeros(N); % Se calculeaza matricea Sw (within class scatter matrix)
for i=1:NumClasses*ImgsPerClass
    sw = sw + msc(:,i) * msc(:,i)';
end
sb = zeros(N); % Se calculeaza matricea Sb (between class scatter matrix)
for i=1:NumClasses
    sb = sb + msm(:,i) * msm(:,i)';
end
```

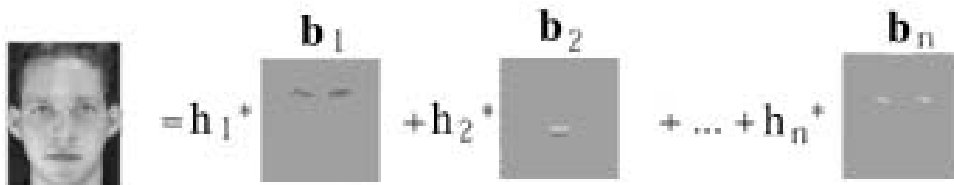
```

% Se rezolva ecuatia: Sb*w = D*Sw*w
[V,D]=eig(sb,sw);
szd = size(D);
for i=1:szd(1)
    evals(i) = D(i,i);
end
[a,b]=sort(evals);
% Se extrag vectorii proprii asociati celor mai mari (NumClasses-1) valori proprii
for i = 1:NumClasses-1
    fisher_basis(:,i)=V(:,b(szd(1)-(i-1)));
end

```

### Tehnici de procesare locală

O serie de lucrări semnificative apărute recent se referă la reprezentarea obiectelor naturale prin părți componente (*parts-based object recognition*), justificată intuitiv de o serie de avantaje potențiale ale acestora precum stabilitatea la deformări locale, grad sporit de invarianță în raport cu nivelul de iluminare, toleranță la acoperire parțială. În cazul particular al procesării fețelor, imaginea se descompune într-o combinație liniară de imagini bine localizate în spațiu, ca în Fig. 1.9, care pot fi interpretate drept componentele unor baze specifice.



**Fig. 1.9** Reprezentarea fețelor prin părți componente

Vom prezenta în continuare câteva modalități diferite de a obține astfel de reprezentări. Procedura generală este următoarea: imaginile folosite sunt reunite sub forma unei matrici  $\mathbf{X}$ , fiecare coloană a matricii reprezentând vectorul intensității luminoase a pixelilor corespunzători unei anumite fotografii. Vom nota cu  $\mathbf{B}$  setul de vectori ai bazei (localizate spațial) și cu  $\mathbf{H}$  matricea coeficienților descompunerii imaginilor în raport cu această bază (acești coeficienți reprezintă proiecții ale matricii de date  $\mathbf{X}$  pe baza  $\mathbf{B}$ ). Dacă numărul de vectori ai bazei este mai mic decât lungimea vectorilor din matricea  $\mathbf{X}$  (și, de regulă, așa se și întâmplă),

atunci se obține compresia datelor de lucru, iar fiecare linie a matricii  $\mathbf{H}$  va forma „semnătura” unei imagini din baza de date. Se poate scrie:

$$\mathbf{X} \approx \mathbf{B}\mathbf{H} \quad (1.23)$$

Procedura de recunoaștere efectivă se va baza pe identificarea distanței minime dintre o astfel de semnătură corespunzătoare unei imagini de test și semnăturile imaginilor din setul de antrenare. Diversele tehnici folosite impun constrângeri specifice asupra lui  $\mathbf{B}$  și/sau  $\mathbf{H}$ , în unele cazuri obținându-se baze localizate spațial.

Metoda NMF a fost recent introdusă ca tehnică liniară de proiecție care impune constrângeri de non-negativitate asupra matricilor  $\mathbf{B}$  și  $\mathbf{H}$  [13]:

$$\mathbf{B}, \mathbf{H} \geq \mathbf{0} \quad (1.24)$$

Justificarea de principiu este legată de metoda intuitivă de combinare a părților pentru a forma un întreg, astfel încât doar combinațiile strict aditive ale vectorilor care formează baza  $\mathbf{B}$  sunt permise. Algoritmul iterativ de calcul pentru obținerea valorilor matricilor de interes este formulat astfel:

$$\begin{aligned} H_{aj} &\leftarrow H_{aj} \sum_i [\mathbf{B}^T]_{ai} \frac{X_{ij}}{[\mathbf{B}\mathbf{H}]_{ij}} \\ B_{ia} &\leftarrow B_{ia} \sum_j \frac{X_{ij}}{[\mathbf{B}\mathbf{H}]_{ij}} [\mathbf{H}^T]_{ja} \\ B_{ia} &\leftarrow \frac{B_{ia}}{\sum_j B_{ja}} \end{aligned} \quad (1.25)$$

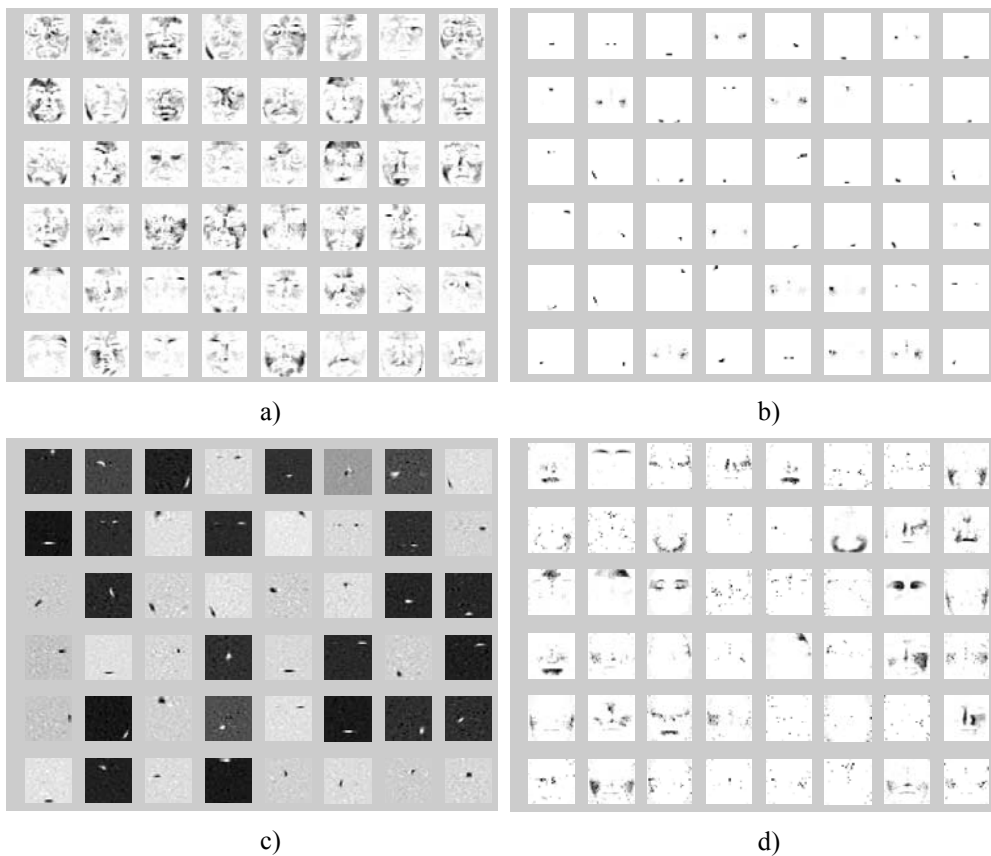
Contrar rezultatelor raportate în articolul original [13], în cazul reprezentării unor imagini de fețe umane care nu sunt în prealabil aliniată cu acuratețe, baza  $\mathbf{B}$  generată prin metoda NMF nu este suficient de bine localizată spațial. Pentru a corecta acest dezavantaj a fost introdusă o versiune locală a algoritmului (denumită *Local NMF* - LNMF), care impune următoarele constrângeri suplimentare [14]:

- a) dispersie maximă a coeficienților matricii  $\mathbf{H}$
- b) maximă expresivitate a vectorilor din baza  $\mathbf{B}$
- c) maximă ortogonalitate pentru vectorii din baza  $\mathbf{B}$

Următoarele ecuații descriu procedura de calcul pentru  $\mathbf{B}$  și  $\mathbf{H}$ :

$$\begin{aligned}
 H_{aj} &\leftarrow \sqrt{H_{aj} \sum_i [\mathbf{B}^T]_{ai} \frac{X_{ij}}{[\mathbf{BH}]_{ij}}} \\
 B_{ia} &\leftarrow B_{ia} \sum_j \frac{X_{ij}}{[\mathbf{BH}]_{ij}} [\mathbf{H}^T]_{ja} \\
 B_{ia} &\leftarrow \frac{B_{ia}}{\sum_j B_{ja}}
 \end{aligned} \tag{1.26}$$

În Fig. 1.10 se prezintă exemple de vectori ai bazei  $\mathbf{B}$  obținuți prin cele două metode descrise anterior, în cazul reprezentării imaginilor din baza de date Olivetti. Se observă că într-adevăr localizarea imaginilor este superioară în cazul algoritmului LNMF.



**Fig. 1.10** Componente ale bazei  $\mathbf{B}$  obținute pe baza algoritmului:  
a) NMF; b) LNMF; c) ICA; d) NA

Este important de menționat că lista metodelor capabile să conducă la baze formate din imagini cu bună localizare spațială este mai largă, incluzând printre altele tehnica ICA (*Independent Components Analysis*) [1], rețelele neurale asociative de tipul celor descrise în contextul PCA, însoțite de constrângerea privind caracterul ne-negativ al tuturor ponderilor (*Non-negative Autoassociators* - NA) [6], sau diversele metode de tip *sparse coding* [7].

### 1.3 Aplicații ale algoritmilor de recunoaștere a fețelor

În cele ce urmează vom prezenta o serie de rezultate experimentale obținute în urma aplicării algoritmilor descriși în paragraful anterior. Vor fi analizate aspecte specifice fiecăruia dintre cele 3 module de bază care compun un sistem automat de recunoaștere a formelor (preprocesare, extragere de trăsături, clasificare), având la dispoziție seturi de imagini disponibile public și utilizate frecvent în literatură pentru analiza comparativă a performanțelor diverselor soluții propuse. Începem prin a descrie sumar bazele de date folosite, vom continua cu prezentarea exemplurilor concrete, iar în final vom enumera o serie de concluzii și comentarii pe marginea rezultatelor obținute. Menționăm că “scenariul” utilizat în fiecare experiment presupune folosirea a 2 seturi de date distincte:

- setul de antrenare (denumit în literatură *gallery set*) reprezintă “materia primă” utilizată de diverși algoritmi pentru a obține informațiile necesare în vederea generării “semnăturii” imaginii analizate. Aceste informații pot fi, de exemplu, vectorii proprii semnificativi ai matricii de covarianță (în cazul algoritmului PCA), sau vectorii care definesc direcțiile de maximă discriminabilitate (în cazul algoritmului LDA).
- setul de test (denumit în literatură *probe set*) este format din imagini “proaspete”, neutilizate în faza de antrenare, care vor servi pentru a estima performanțele reale ale metodei de recunoaștere studiate.

Pentru ca rezultatele să fie relevante din punct de vedere statistic, de obicei se efectuează *experimente repetate*, alocând în mod aleator imaginile din baza de date disponibilă către setul de antrenare, respectiv cel de test, iar valorile ratelor de

recunoaștere rezultate în fiecare experiment *se mediază* pentru a obține în final un rezultat care *estimează* nivelul real de performanță al soluției analizate.

#### *Baze de date utilizate în aplicații de recunoaștere a fețelor*

În literatură sunt raportate performanțe de recunoaștere utilizând o listă destul de largă de baze de date, majoritatea disponibile gratis în scopuri de cercetare. Acestea diferă prin numărul persoanelor incluse, numărul de fotografii ale aceluiași subiect, rezoluție și format, iar dintre acestea amintim [5]:

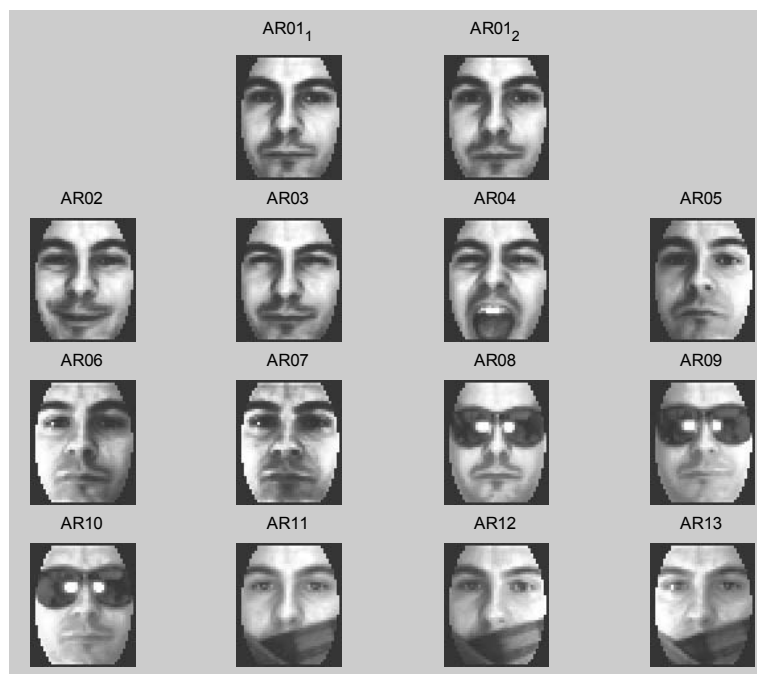
- *Baza de date Olivetti*: conține câte 10 imagini distincte pentru un număr total de 40 de persoane, prezentând variații din punctul de vedere al orientării, rotației în planul imaginii (de până la 20°), scalei de reprezentare (de până la 10%) și expresiei feței, în condiții de iluminare controlate. Dimensiunile fiecărei imagini sunt de 112x92 pixeli, folosind 256 nivele de gri, iar exemple se prezintă în Fig. 1.11. În experimentele efectuate am alocat în mod aleator un număr de maxim 5 imagini (din totalul de 10) ale fiecărei persoane setului de antrenare, iar restul au format setul de test.
- *Baza de date AR*: conține 113 persoane (63 bărbați și 53 femei), fotografiile având 768x576 pixeli, reprezentate color pe 24 de biți. Pentru fiecare subiect sunt incluse 2 seturi de câte 13 imagini, obținute la interval de două săptămâni, caracterizate de expresie variabilă a feței, nivele diferite de iluminare și acoperire parțială, ca în exemplele din Fig. 1.12. Conform metodologiei utilizate frecvent în literatură, am alocat setului de antrenare numai perechile de imagini corespunzătoare expresiei neutre (notate AR0<sub>1</sub>, respectiv AR0<sub>2</sub> în Fig. 1.12), urmând ca setul de test să fie format din restul imaginilor disponibile. În particular, imaginile au fost preprocesate prin mascarea fundalului și a părului<sup>1</sup>.
- *Baza de date Yale*: conține 165 de imagini ale unui număr de 15 persoane, caracterizate prin variabilitate semnificativă a expresiei feței, ca în Fig. 1.13. Datorită dimensiunii mici a bazei de date, tipic se efectuează experimente repetate alocându-se setului de test numai câte o singură imagine a fiecărei persoane.

---

<sup>1</sup> Baza de date preprocesată a fost pusă la dispoziție cu amabilitate de către David Guillaumet de la Universitat Autònoma de Barcelona.



**Fig. 1.11** Exemple de imagini din baza de date Olivetti



**Fig. 1.12** Exemple de imagini din baza de date AR



**Fig. 1.13** Exemple de imagini din baza de date Yale



*Tehnici de preprocesare*

Multe dintre tehnicile de recunoaștere a fețelor, în particular metoda PCA, sunt foarte sensibile la alinierea corectă a imaginilor care formează setul de antrenare, performanțele degradându-se semnificativ dacă acestea sunt ușor translate sau rotite. Din acest motiv, multe soluții folosesc o etapă premergătoare de preprocesare (automată sau manuală) care are drept rezultat obținerea unei baze de date formate din imagini având aceeași scală de reprezentare și neafectate de translații relative sau rotații în planul imaginii (tipic, segmentul definit de poziția ochilor este folosit pentru această etapă de normalizare).

În plus, există argumente teoretice care justifică utilizarea exclusivă a zonei care definește fața propriu-zisă, eliminându-se influența fundalului și a părului. În acest scop se utilizează “măști” de formă ovală, precum se indică în Fig.1.12 în cazul bazei de date AR.

Dimensiunile mari ale imaginilor reprezintă de multe ori o problemă majoră, deoarece volumul de calcul poate deveni excesiv în unele cazuri. Ținând cont de rezultate care confirmă degradarea lentă a performanțelor de recunoaștere în raport cu rezoluția folosită, se pot aplica proceduri de subeșantionare care să conducă la dimensiuni rezonabile (de exemplu, de ordinul a 1000-2000 de pixeli). În particular, se poate face apel la Transformata Wavelet Discretă (prezentată pe larg în capitolul următor), care permite reducerea dimensionalității simultan cu asigurarea invarianței în raport cu expresia feței.

*Metoda PCA și variantele sale*

Metoda PCA (denumită *eigenfaces* în contextul aplicațiilor de recunoaștere a fețelor [24]) reprezintă fără îndoială referința în raport cu care se compară performanțele oricărui algoritm propus în literatură. Deși simplu de aplicat și oferind performanțe rezonabile în condiții dintre cele mai diverse, metoda este totuși afectată de unele limitări precum sensibilitatea în raport cu alinierea imaginilor și nivelul de iluminare, cerința de a beneficia de o bază de date de dimensiune suficient de mare, dificultatea de a opera *on-line* (în sensul de a putea actualiza fără un efort de calcul substanțial valorile vectorilor proprii semnificativi odată cu apariția unor date „proaspete” în setul de antrenare). Ca urmare, au apărut

o serie de variante ale algoritmului standard, care încearcă să diminueze aceste neajunsuri.

Astfel, îmbunătățirea toleranței în raport cu alinierea imaginilor este sugerată prin metoda denumită (PC)<sup>2</sup>A (*Projection Combined PCA*) [26], care definește mai întâi așa-numitele proiecții integrale de-a lungul axei orizontale și verticale ale unei imagini  $I(x, y)$  de dimensiune  $M \times N$  prin:

$$\mathbf{V}_p(x) = \sum_{y=1}^N \mathbf{I}(x, y), \quad \mathbf{H}_p(y) = \sum_{x=1}^M \mathbf{I}(x, y) \quad (1.27)$$

pe baza cărora se definește apoi matricea:

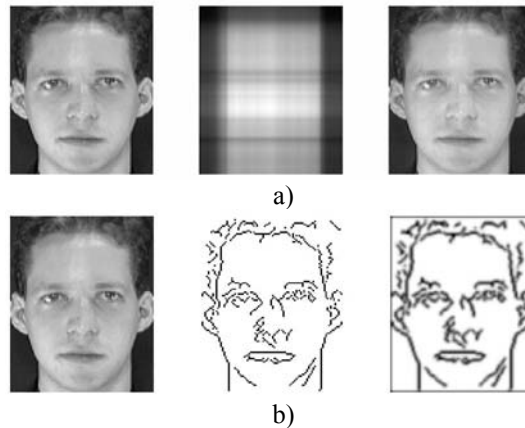
$$\mathbf{M}_I(x, y) = \frac{\mathbf{V}_p(x)\mathbf{H}_p(y)}{MN\bar{\mathbf{I}}}, \quad \bar{\mathbf{I}} = \frac{\sum_{x=1}^M \sum_{y=1}^N \mathbf{I}(x, y)}{MN} \quad (1.28)$$

În final se obține o versiune preprocesată a imaginii originale conform relației:

$$\mathbf{P}_\alpha(x, y) = \frac{\mathbf{I}(x, y) + \alpha \mathbf{M}_I(x, y)}{1 + \alpha} \quad (1.29)$$

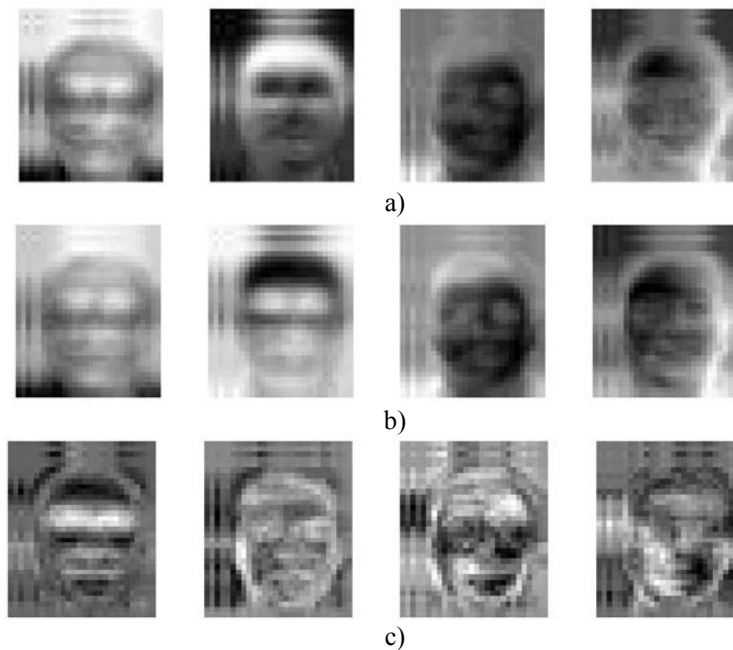
Efectul intuitiv al acestei proceduri constă într-un anumit grad de dezaccentuare a conturilor ferme ale imaginii originale (*blurring*), după cum se ilustrează în Fig. 1.14a. Utilizarea efectivă a metodei presupune aplicarea acestei transformări asupra fiecărei imagini din baza de date, urmând ca apoi să folosim varianta standard a algoritmului PCA.

O altă soluție își propune să atenueze sensibilitatea PCA în raport cu nivelul variabil de iluminare. Astfel, pornind de la observația intuitivă potrivit căreia iluminarea influențează destul de puțin conturile unei imagini, metoda *eigenhills* [27] propune detecția prealabilă a conturilor fețelor din baza de date, după care aplică metoda PCA standard. În realitate, pentru a preîntâmpina dezavantajul conform căruia imaginile reprezentând conturile sunt mult mai sensibile la translații relative și expresii variabile ale feței, se aplică asupra conturilor o operație de filtrare trece-jos, astfel încât informația este „împrăștiată” oarecum în spațiu, după cum se sugerează în Fig.1.14b.



**Fig. 1.14** Variante ale algoritmului PCA : a) metoda  $(PC)^2A$  ; b) metoda *eigenhills*

Exemple de componente ale bazei rezultate din aplicarea fiecărei variante se prezintă în Fig. 1.15. Rezultate experimentale comparative între cele 3 metode folosind baza de date Olivetti sunt indicate în Tabelul 1.1. Au fost efectuate 10 experimente distincte, folosind câte 5 imagini ale fiecărei persoane în setul de antrenare, respectiv de test.



**Fig. 1.15** Exemple de componente ale bazelor obținute prin metoda :  
a) PCA standard; b)  $(PC)^2A$  ; c) *eigenhills*

Tabelul 1.1: Rezultate experimentale PCA,  $(PC)^2A$ , *eigenhills*

Tipul distanței	Eigenface	$(PC)^2A$	Eigenhills
L2 (individual)	94.7 %	95.5 %	84.4 %
L2 (average)	90.8 %	89.6 %	82.5 %
Mahalanobis	94.8 %	95.4 %	85.4 %

*Tehnici de procesare locale*

În Fig. 1.9 a fost ilustrată posibilitatea de a reprezenta fețele prin părți componente sau, mai exact, de a încerca descompunerea unei astfel de imagini în raport cu o bază particulară, ale cărei componente să fie caracterizate de o bună localizare spațială. Drept exemple de algoritmi capabili să conducă la obținerea unei astfel de baze au fost menționate metodele NMF (*Non-negative Matrix Factorization*), varianta locală a acesteia LNMF (*Local NMF*), algoritmul ICA (*Independent Components Analysis*) și rețelele neurale autoasociative cu parametri nenegativi NA (*Non-negative Autoassociators*), iar exemple de componente ale acestor baze sunt prezentate în Fig. 1.10.

Utilizând baza de date AR au fost efectuate o serie de experimente care își propun să analizeze comparativ performanțele acestor algoritmi în raport cu diverse surse de variabilitate precum expresia feței și nivelul de iluminare, iar rezultatele se prezintă în Tabelele 1.2 și 1.3 (parametrul  $m$  indică numărul vectorilor care formează baza, adică dimensiunea subspațiului pe care se efectuează proiecția vectorilor care reprezintă imaginile originale). Aceleleași metode au fost utilizate și pentru studierea performanțelor în situațiile în care fețele sunt afectate de acoperire parțială. Rezultatele corespunzătoare bazei AR sunt indicate în Tabelul 1.4, observându-se o degradare semnificativă a performanțelor (sunt prezentate numai rezultate referitoare la cazul acoperirii folosind ochelari, deoarece ochii au o importanță specială în stabilirea nivelului de performanță al unui sistem automat de recunoaștere a fețelor). În cazul bazei de date Olivetti acoperirea a fost modelată prin suprapunerea unor măști pătrate de dimensiune variabilă, plasate aleator în planul imaginii, după cum se sugerează în Fig. 1.16, iar rezultatele sunt prezentate în Fig. 1.17.

Tabelul 1.2: Rate de recunoaștere pentru expresie variabilă (AR)

	AR02				AR03				AR04			
	m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200
L2	83.7	72.2	64.5	82.4	89.7	92.3	93.1	95.3	41.4	46.1	39.7	49.5
L1	92.7	92.3	86.3	95.7	93.1	94	94.8	96.5	53	56.8	54.7	61.5
LNMF + cos	76	64.1	59.4	73.9	84.6	90.6	92.7	93.6	34.2	38.9	33.3	41.8
L2	91	92.3	92.7	91.8	91.4	92.3	93.1	93.6	49.1	52.1	53	55.1
ICA + L1	91	92.7	91.8	91.4	93.1	93.6	93.6	94.4	51.7	55.1	55.5	57.2
cos	89.7	90.6	91.4	89.7	89.3	90.6	91	90.1	58.1	62.4	60.6	61.1
L2	79	91	92.7	93.1	67.9	85.9	89.7	88.4	29.5	38.9	38.9	44.4
NMFsc + L1	88.9	95.7	96.1	93.6	86.7	91.8	92.7	90.6	41.8	44	46.5	46.5
cos	73.5	88	91.8	91.8	65.8	85.4	91	89.3	26.9	37.1	38.9	45.7
PCA	91	94.4	95.3	95.7	88	89.7	89.7	90.6	47.4	52.5	52.5	52.5

Tabelul 1.3: Rate de recunoaștere pentru iluminare variabilă (AR)

	AR05				AR06				AR07			
	m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200
L2	17	29.5	36.3	25.6	11.9	13.6	11.9	8.9	2.1	6.4	3.4	1.7
L1	20	32.9	38.4	28.2	8.1	20	13.6	11.5	1.2	1.7	2.1	2.1
LNMF + cos	46.5	48.3	53.8	57.2	30.7	23	20.9	10.2	17	15.3	14.5	17
L2	95.3	97.4	97.4	98.3	89.3	92.7	93.6	93.1	73.9	79.5	80.3	79.9
ICA + L1	95.3	97	97.8	97.4	90.6	92.3	94.8	92.7	75.6	79	79.5	79
cos	95.7	97.4	97.4	97	94	97.4	97.8	97.4	88.4	89.3	89.3	88.9
L2	44	56	76	71.3	9.8	22.6	22.2	34.2	11.1	15.3	23.5	27.3
NMFsc + L1	43.1	53	73	76	9.4	26.5	17.9	32	5.1	10.6	19.6	20.9
cos	55.1	61.9	77.3	73.9	11.9	27.7	25.6	36.7	22.6	24.3	34.6	37.6
PCA	73.5	77.3	80.7	81.2	16.2	20.9	21.3	21.3	58.9	67	70	71.3



Fig. 1.16 Imagini din baza de date Olivetti acoperite parțial

Tabelul 1.4: Rate de recunoaștere pentru acoperire parțială (AR)

		AR08				AR09				AR10			
		m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200	m=50	m=100	m=150	m=200
LNMF+ cos	L2	7.7	6.8	5.5	7.2	5.1	5.5	3.4	3.8	5.1	2.5	3.8	3.8
	L1	22.2	20.5	17	28.6	10.6	14.1	12.8	14.1	9	6.4	5.1	7.7
	cos	8.1	6	2.5	6.8	4.2	4.7	3	3.8	4.7	2.1	2.1	3
ICA + L1	L2	28.2	34.6	34.6	35.9	26	27.7	29	29.5	26	29	29.5	30.3
	L1	26.9	29.5	30.7	32.9	27.3	25.2	26	28.6	26.5	25.6	27.3	27.3
	cos	39.3	43.6	45.3	47.4	36.3	38.9	40.6	40.6	31.6	36.7	36.7	38
NMF <sub>sc</sub> + L1	L2	10.2	9.8	11.5	17	5.5	8.1	7.7	11.5	6.4	6.8	8.1	8.5
	L1	18.3	14.5	15.3	23.9	9.4	9.4	8.1	11.1	5.5	6.4	7.7	9.4
	cos	9.8	9.4	9.8	18.3	4.2	7.2	7.2	11.1	5.1	6	6.4	7.7
PCA		8.5	8.5	10.2	11.1	11.5	12.4	13.2	13.2	8.9	9.8	9.4	9.8

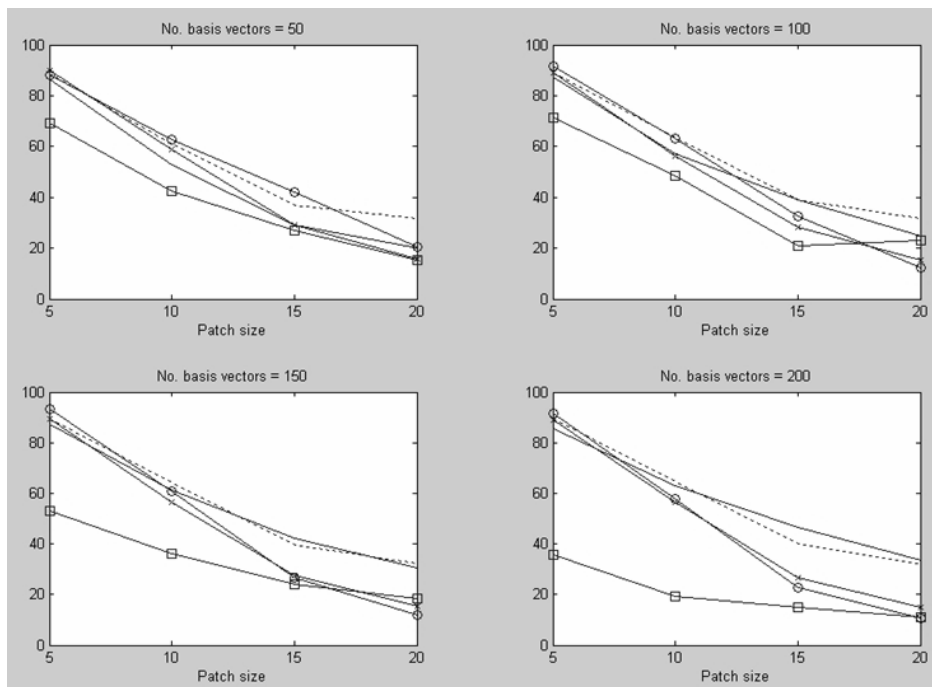


Fig. 1.17 Rate de recunoaștere pentru baza de date Olivetti

### Clasificare folosind rețele neurale autoasociative

Principiul de clasificare denumit generic “regula celui mai apropiat vecin” (*nearest neighbor rule*) rămâne în continuare soluția cea mai des utilizată în aplicații generale de recunoaștere de forme, în particular pentru problemele orientate pe

prelucrarea imaginilor reprezentând fețe umane. Deși în literatură au fost propuse și alternative mai performante (de exemplu, bazate pe utilizarea clasificatoarelor optimale de tip Bayes), simplitatea rămâne un argument puternic în favoarea primei metode. O alternativă interesantă este oferită de teoria sistemelor dinamice neliniare, în mod concret de posibilitatea de a sintetiza așa-numite *memorii asociative* [4], capabile să furnizeze ieșirea solicitată utilizând drept „cheie” nu o adresă, așa cum se întâmplă în cazul memoriilor clasice, ci o replică, de regulă zgomotoasă, incompletă sau eronată, a însăși informației stocate (acesta este și motivul pentru care astfel de sisteme sunt denumite și *memorii adresabile prin conținut*). Ideea fundamentală este următoarea: trebuie sintetizat un sistem dinamic care să admită un set de *puncte de echilibru stabil* în dreptul vectorilor care urmează a fi stocați (de exemplu, în aplicațiile de recunoaștere a fețelor, vectorii memorati pot reprezenta „semnăturile” corespunzătoare imaginilor din setul de antrenare). Un vector zgomotos și/sau distorsionat (utilizat drept „cheie” pentru recuperarea informației de interes; de exemplu, „semnăturile” fețelor din setul de test) va reprezenta în fapt *o stare inițială* din care sistemul va evolua în mod natural spre starea de echilibru stabil în *bazinul de atracție* al căreia se află plasată „cheia”, în mod ideal către cel mai apropiat în distanță Euclideană. Ar fi de dorit ca asemenea sisteme să ofere următoarele caracteristici:

- să admită puncte de echilibru numai în poziții predefinite, fără apariția unor parazite
- numărul de stări de echilibru stabil trebuie să fie în principiu oricât de mare
- introducerea sau eliminarea unui punct de echilibru trebuie să se facă simplu, fără a reproiecta întregul sistem
- extensia bazinului de atracție în jurul punctelor de echilibru trebuie să fie controlabilă

O posibilă soluție de sinteză a unui astfel de sistem dinamic (neliniar) este reprezentat de un sistem *de tip gradient*, definit prin relația:

$$\frac{dx_i}{dt} = -\frac{\partial V(\mathbf{X})}{\partial x_i}, \quad \mathbf{X} = [x_1 \ x_2 \ \dots \ x_N]^T \quad (1.30)$$

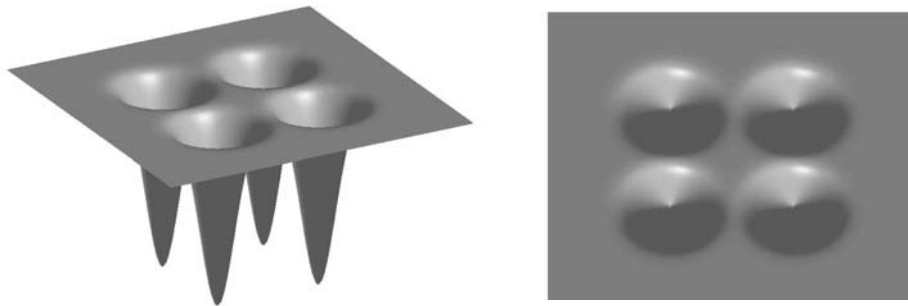
unde  $N$  reprezintă dimensiunea vectorului de stare, iar  $V(\mathbf{X})$  desemnează așa-numita **funcție Lyapunov**. Un rezultat teoretic binecunoscut afirmă că punctele de echilibru stabil ale unui sistem de tip gradient coincid cu *minimele izolate* ale funcției Liapunov. Ținând cont de aceasta, cerințele formulate anterior pot fi satisfăcute definind funcția Liapunov într-o manieră particulară, sub forma unei *sume de funcții individuale cu selectivitate spațială*, fiecare având un minim accentuat în dreptul unui singur punct și fiind practic constante în rest [4]:

$$V(\mathbf{X}) = \sum_{s=1}^M g_s(\mathbf{X}) \quad (1.31)$$

O posibilitate constă în a alege drept argument al funcției selective  $g_s(\cdot)$  distanța (convenabil aleasă) dintre vectorul de stare curent  $\mathbf{X}$  și un vector reprezentând poziția dorită a unei informații ce se dorește a fi stocată, astfel încât vom utiliza funcții selective prezentând un minim accentuat în origine:

$$V(\mathbf{X}) = \sum_{s=1}^M g_s(\mathbf{X}) = \sum_{s=1}^M G_s[d(\mathbf{X}, \mathbf{X}^s)] \quad (1.32)$$

Un exemplu concret îl reprezintă alegerea unei funcții de tip gaussian, recomandată de considerente legate de facilitățile de implementare pe care le oferă, axate pe posibilitatea de factorizare comodă. O imagine intuitivă asupra funcției Liapunov definite mai sus se prezintă în Fig. 1.18 ( $N=2$ ,  $M=4$ ; punctele de echilibru sunt:  $(-1,-1)$ ,  $(-1,1)$ ,  $(1,-1)$ ,  $(1,1)$ , iar  $\sigma_s=1$ ).



**Fig. 1.18** Exemplu de funcție Liapunov de tipul (1.32)



Procedura de clasificare oferă avantaje evidente față de principiul celui mai apropiat vecin: nu este necesar să evaluăm distanța dintre vectorul de test și toți vectorii care formează setul de antrenare (fapt care ar conduce la un volum mare de calcul atunci când baza de date are dimensiune mare), ci este suficient să inițializăm starea sistemului dinamic neliniar cu vectorul de test, iar evoluția acestuia către unul dintre punctele de echilibru ne va furniza de fapt valoarea vectorului de antrenare cel mai apropiat de cel de test. Soluția prezintă un grad sporit de modularitate deoarece introducerea/eliminarea unui punct de echilibru nu va influența semnificativ poziția celorlalte puncte, oferă o corespondență clară între setul de informații ce urmează a fi stocate și ecuațiile care guvernează dinamica sistemului, efectul diversilor parametri asupra evoluției temporale este ușor de interpretat, iar implementarea beneficiază de avantajul unui set redus de interconexiuni.

Am utilizat principiul descris anterior în cazul bazei de date Olivetti, folosind drept vectori ce urmează a fi memorați „semnături” extrase pe baza algoritmilor PCA, respectiv LDA (în realitate, aplicarea algoritmului LDA a fost precedată de utilizarea metodei PCA în vederea reducerii dimensionalității datelor), iar rezultatele se prezintă în Tabelul 1.5. Au fost efectuate 10 experimente distincte, folosind câte 5 imagini ale fiecărei persoane pentru setul de antrenare, iar dimensiunea vectorilor de trăsături extrase a variat între 50 și 200 (în tabel sunt prezentate performanțele optime pentru fiecare categorie de „semnături”). Este ușor de observat că metoda conduce la performanțe comparabile cu cele raportate în literatură folosind alte soluții.

*Tabelul 1.5:* Rezultate comparative folosind baza de date Olivetti

<b>Metoda</b>	<b>Eroare de recunoaștere (%)</b>
<i>Eigenfaces</i>	10
<i>Convolutional Neural Network</i>	3.8
<i>Linear SVM</i>	3
<i>Kernel PCA</i>	2.5
PCA + Memorie asociativă	5.4
LDA + Memorie asociativă	3.1

#### 1.4 Discuție asupra sistemelor biometrice

Prezentarea din cuprinsul paragrafelor anterioare s-a concentrat asupra aplicațiilor orientate pe procesarea imaginilor reprezentând fețe umane, ilustrate pe baza unor exemple care să ofere o imagine intuitivă asupra utilizării diversilor algoritmi enumerați. Extinzând cadrul discuției la ansamblul tehnologiilor biometrice, merită trecute în revistă o serie de elemente specifice acestui domeniu, care trebuie avute în vedere în cazul implementării unor sisteme automate capabile să furnizeze nivele satisfăcătoare de performanță în aplicații realiste.

##### *Tehnologii biometrice*

Există o listă destul de lungă a tipurilor de informații biometrice care sunt utilizate în mod curent în aplicații de recunoaștere sau autentificare, care include printre altele: vocea, amprentele, fața, irisul, forma geometrică a mâinii, stilul de scriere, alura mersului, precum și combinații ale acestora. Sistemele comerciale folosesc de regulă primele patru categorii, datorită existenței senzorilor necesari cu prețuri și performanțe satisfăcătoare, dar și a suportului teoretic aferent procesării adecvate a datelor disponibile. În continuare vom enumera avantajele și dezavantajele celor mai importante tehnologii:

- *ampretele*: au avantajul că se păstrează practic nealterate în decursul vieții și și-au demonstrat deja eficiența în practică, existând un volum apreciabil de date colectate. Prețul suficient de scăzut al multor senzori, diversitatea tehnologiilor de obținere și dimensiunile reduse sunt de asemenea atractive. Ca dezavantaje putem menționa posibilitatea ca anumite profesii să poată conduce la alterarea calității amprentelor, obținerea unor date de bună calitate poate necesita o anumită procedură de „antrenare”, precum și reticența de a atinge un senzor pe care l-au atins în prealabil foarte multe persoane.
- *irisul*: nu necesită un contact fizic cu senzorul și reprezintă o informație extrem de individualizată (probabilitatea ca două persoane diferite să aibă o aceeași „semnătură” biometrică extrasă pe baza irisului este de  $3 \times 10^{-8}$ !). Pe de altă parte, există o rețineră în a expune ochii către o sursă de lumină, este necesară o pricepere mai mare decât în alte cazuri pentru a extrage informația de interes, există riscul ca

informația utilă să fie mascată de prezența unor lentile, a genelor sau a reflexiilor de pe cornee. Tehnologia de procesare a unor astfel de informații a fost elaborată la începutul anilor '80 de către cercetătorul american John Daugman și este protejată de o serie de patente.

- *fața*: reprezintă de asemenea o tehnică neinvazivă, utilizează drept senzori camere de fotografiat sau filmat disponibile pe scară largă, există baze de date de dimensiune mare, iar rezultatele pot fi verificate cu ușurință. Pe de altă parte, fețele sunt puternic afectate de nivelul de iluminare, expresie, transformări geometrice, îmbătrânire, iar unele aplicații video presupun viteză de operare sporită. În practică, pe lângă distincția dintre problemele de recunoaștere și cele de verificare, se disting câteva direcții particulare de folosire a acestor informații precum analiza expresiei feței, determinarea stării emoționale, detecția și urmărirea ochilor sau a buzelor.
- *vocea*: este acceptată pe scară largă, senzorii sunt la îndemână și nu necesită contact fizic cu utilizatorul. În schimb, vocea prezintă variații semnificative în raport cu vârsta, starea de sănătate și cea emoțională, apar dificultăți datorate alterării vocii de către tipul microfonului utilizat și a zgomotului ambiental, iar rezultate satisfăcătoare sunt dificil de obținut pe baze de date de dimensiune mare. Sunt avute în vedere două familii mari de aplicații, de recunoaștere a vorbirii (indiferent de persoana care rostește cuvintele, precum în cazul roboților telefonici), respectiv de identificare a vorbitorului.

#### *Proceduri de evaluare a performanțelor*

Înainte ca un sistem biometric automat să devină pe deplin operațional, acesta este supus unui ansamblu de teste specifice, fiecare având un scop precis, fixat în acord cu o modalitate particulară de analiză și utilizând o bază de date adecvată. Putem identifica următoarele tipuri de proceduri de evaluare [18]:

- *evaluarea tehnologiei*: urmărește validarea metodei analizate folosind baze de date standardizate de dimensiune mică, identificând aspectele critice care necesită o analiză teoretică suplimentară și furnizând nivele de performanță cu rol de referință pentru celelalte tipuri de proceduri de testare.
- *evaluarea pe bază de scenariu*: durează mai mult ca în cazul precedent și își propune să măsoare performanțele pentru o aplicație specifică, folosind baze de

date de dimensiuni mari, achiziționate de la un număr de utilizatori nefamiliarizați în prealabil cu sistemul. Rezultatele pot evidenția zone care necesită analiză suplimentară sau posibilități de integrare și oferă valori mai exacte ale nivelelor reale de performanță.

▪ *evaluarea operațională*: are în vedere cuantificarea impactului pe care introducerea sistemului biometric îl are asupra ansamblului în care acesta este inclus. În acest context, este necesară evaluarea stării existente înainte de introducerea sistemului, respectiv celei de după familiarizarea utilizatorilor cu modalitatea concretă de operare a acestuia. Astfel de teste pot avea durate mari, de ordinul câtorva luni, iar aspectele vizate se referă atât la tehnologia propriu-zisă, cât și la efectele psihologice și economice pe care aceasta le produce.

#### *Modalități de fraudare*

Nu există sisteme biometrice ideale: implementarea oricărei metode de recunoaștere/verificare bazate pe utilizarea unor astfel de informații este inevitabil limitată inferior de valori a căror estimare precisă reprezintă subiectul unor proceduri riguroase de evaluare standardizată și/sau al unor competiții de mare anvergură, precum cele organizate periodic de către organizația guvernamentală americană *National Institute of Standards and Technology* ([www.nist.gov](http://www.nist.gov)). În particular, aplicațiile de verificare (autentificare) trebuie să asigure un compromis rezonabil între cele două tipuri majore de erori (rata de acceptare, respectiv de rejecție falsă), astfel încât să minimizeze probabilitatea de acces la resurse/spații protejate a unor persoane neautorizate fără a deranja excesiv pe utilizatorii autorizați. Alegerea concretă a unor valori limită pentru aceste erori depinde nemijlocit de aplicația considerată și de restricțiile aferente, astfel încât în realitate vor exista întotdeauna “portite” care pot fi folosite în tentative de fraudare.

Pe de altă parte, folosind seturi de date generate *artificial*, au fost raportate și demonstrate posibilități de a “păcăli” destul de ușor chiar și sisteme comerciale care s-au dovedit performante în testele de evaluare. Spre exemplu, cercetătorul japonez Tsutomu Matsumoto și grupul său de la Yokohama National University au obținut amprente artificiale pe bază de gelatină, pornind atât de la amprentarea directă a unor degete, cât și de la amprente *latente*, prelevate de pe o bucată de

sticlă. În ambele variante, o serie de sisteme comerciale au fost „păcălite” în 80% dintre cazuri!

În cazul aplicațiilor de autentificare a fețelor, un studiu recent publicat în revista germană *c't* a arătat că unul dintre sistemele cele mai bune de pe piață, FaceVACS produs de către firma Cognitec (una dintre câștigătoarele detașate ale ultimei competiții *Face Recognition Vendor Test!*), poate fi fraudat prin prezentarea în fața sensorului optic a unei fotografii corespunzătoare unei persoane autorizate sau rularea unui scurt fișier video. Ca urmare, firma a introdus o procedură suplimentară de detecție a prezenței fizice a unei persoane în fața sensorului (*Live-Check*), însă acest element a îngreunat accesul utilizatorilor autorizați. De altfel, rularea unei înregistrări în care o persoană își mișcă ușor capul dintr-o parte în alta a „păcălit” din nou sistemul.

Nici în cazul recunoașterii irisului situația nu este cu mult mai bună, deși această informație biometrică este recunoscută ca fiind cea mai puțin repetabilă. „Portița” de intrare o reprezintă particularitatea că multe sisteme comerciale prelevează în realitate nu numai porțiunea strict ocupată de iris, ci și o zonă de dimensiune variabilă din jurul acestuia. Astfel, jurnaliștii de la aceeași revistă germană au efectuat un test cu sistemul Authenticam BM-ET100 produs de către firma Panasonic, folosind fotografia ochiului unei persoane autorizate, dar decupată în centru pentru a permite inspectarea irisului unei alte persoane, iar sistemul nu a sesizat diferența! Pe de altă parte însă, proiectarea *imaginii digitizate* a irisului unei persoane autorizate de pe ecranul unui *notebook* nu a reușit să păcălească sistemul.

#### *Etica profesională*

Potențialul unei tehnologii biometrice trebuie judecat nu numai din punct de vedere tehnic, dar și din perspectiva unor potențiale utilizări abuzive. Există o serie de decizii importante care trebuie luate înainte de introducerea în exploatare a unui astfel de sistem, printre care evidențierea nivelului satisfăcător de performanță, identificarea informațiilor private care sunt strict necesare, estimarea raportului dintre costuri și sporul de securitate pe care introducerea unei astfel de tehnologii îl poate aduce. Răspunsul la aceste cerințe trebuie să se bazeze pe un anumit cod de etică profesională, care să stipuleze explicit preocuparea proiectanților pentru

aspecte ce țin de protejarea datelor cu caracter personal și respectarea drepturilor omului, obligația acestora de a furniza persoanelor cu putere de decizie, dar și publicului larg, date reale și complete despre performanțele, constrângerile și limitele sistemului analizat, cu scopul de a contribui în cele din urmă la îmbunătățirea calității vieții și a siguranței utilizatorilor. De exemplu, în Marea Britanie au fost dezbateri aprinse pe marginea implementării unui sistem ce urmează să stocheze informații de natură biometrică ale întregii populații, cu formularea unor temeri cu privire la fiabilitatea unei baze de date de asemenea dimensiuni și la consecințele critice ale compromiterii confidențialității acestora. Unul dintre elementele cheie ale eticii profesionale îl reprezintă aprecierea obiectivă a performanțelor unor astfel de sisteme, fără a le supraestima într-un mod nerealist numai din considerente de noutate a unui subiect aflat actualmente „la modă”, iar din acest punct de vedere menționăm următoarele comentarii:

*"As methods of identification, however, biometric technologies are still immature, and one, face recognition, has been especially disappointing. In a test this spring of a leading system, that of Jersey City, N.J.-based Visionics Corp. (now merged with Identix Inc., Minnetonka, Minn.), over half the faces in a mock terrorist database used at the Palm Beach (Fla.) International Airport were let through unflagged, while one person in every hundred to pass through the system was falsely labeled "terrorist" "*

Steven Cherry, *Senior Associate Editor, IEEE Spectrum*, Septembrie 2002

*"The electronic passport puts up a Maginot line at the border, when what we really need is a comprehensive defense that impedes the aspiring terrorist — but not innocent travelers — at every step."*

Philip E. Ross, *IEEE Spectrum*, Ianuarie 2005

---

## Bibliografie

- [1] Bartlett, M.S., Movellan, J.R., Sejnowski, T.R., "Face recognition by Independent Component Analysis", *IEEE Trans. on Neural Networks*, 13, pp. 1450-1464, 2002
- [2] Belhumeur, P.N., Hespana, J.P., Kriegman, D.J., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Trans. PAMI*, 19, no. 7, pp. 711-720, 1997
- [3] Belhumeur, P.N., Kriegman, D.J., "What is the Set of Images of an Object under All Possible Illumination Conditions?", *Int. J. Computer Vision*, vol. 28, no. 3, pp. 1-16, 1998
- [4] Ciocoiu, I.B., *Rețele neurale artificiale*, Editura Cantes, Iași, 2001
- [5] *Face Recognition Homepage*: [www.face-rec.org/databases](http://www.face-rec.org/databases)
- [6] Ge, X., Iwata, S., "Learning the parts of objects by auto-association", *Neural Networks*, vol. 15, pp. 285-295, 2002
- [7] Hoyer, P.O., "Non-negative matrix factorization with sparseness constraints", *J. Machine Learning Research*, vol. 5, pp. 1457-1469, 2004
- [8] *IEEE Spectrum*, vol. 41, no. 3, pp. 13, 2004
- [9] Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A., "Recent advances in visual and infrared face recognition—a review", *Computer Vision Image Understanding*, vol. 97, pp. 103-135, 2005
- [10] Kreutz, M., Volpel, B., Jansen, H., "Scale-Invariant Image Recognition Based on Higher Order Autocorrelation Features", *Pattern Recognition*, vol. 29, no. 1, pp. 19-26, 1996
- [11] Kurita, T., Hotta, K., Mishima, T., "Scale and rotation invariant recognition method using high-order local autocorrelation features of log-polar image", *Proc. Asian Conf. on Computer Vision*, pp. 89-96, 1998
- [12] Lai, J.H., Yuen, P.C., Feng, G.C., "Face Recognition Using Holistic Fourier Invariant Features", *Pattern Recognition*, vol. 34, pp. 95-109, 2001
- [13] Lee, D.D., Seung, H.S., "Learning the parts of objects by non-negative matrix factorization", *Nature*, 401, pp. 788-791, 1999
- [14] Li, S.Z., Hou, X.W., Zhang, H.J., "Learning spatially localized, parts-based representation", *Proc. CVPR*, 2001

- [15] Liu, X., Chen, T., Vijaya Kumar, B.V.K., "Face Authentication for Multiple Subjects Using Eigenflow", *CMU Technical Report AMP 01-05*, 2001
- [16] Liu, C., Wechsler, H., "Independent component analysis of Gabor features for face recognition", *IEEE Trans. Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003
- [17] Moghaddam, B. , "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", *IEEE Trans. PAMI*, vol. 24, no.6, 2002
- [18] *NIST Subcommittee on Biometrics*: [www.biometricscatalog.org](http://www.biometricscatalog.org)
- [19] Penev, P., Atick, J., "Local feature analysis: A general statistical theory for object representation", *Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477-500, 1996
- [20] Pentland, A., Moghaddam, B., Starner, T., "View-based and modular eigenspaces for face recognition", *Proc. CVPR*, pp. 84-91, 1994
- [21] Piotrowski, L.N., Campbell, F.W., "A demonstration of the visual importance and flexibility of spatial-frequency, amplitude, and phase", *Perception*, vol. 11, pp. 337-346, 1982
- [22] Shan, S., Chang, Y., Gao, W., Cao, B., "Curse Of Mis-Alignment In Face Recognition: Problem And A Novel Mis-Alignment Learning Solution", *Proc. AFGR*, pp. 314-320, 2004
- [23] Simard, P.Y., Le Cun, Y.A., Denker, J.S., Victorri, B., "Transformation Invariance in Pattern Recognition – Tangent Distance and Tangent Propagation", *Int. J. Imaging System and Technology*, vol. 11, Issue 3, pp. 181-194, 2001
- [24] Turk, M., Pentland, A.P., "Eigenfaces for recognition", *J. of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991
- [25] Wiskott, L., Fellous, J.-M., Kruger, N., von der Malsburg, C., "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans. PAMI*, vol. 17, no. 7. pp. 775-779, 1997
- [26] Wu, J., Zhou, Z.H., "Face recognition with one training image per person", *Pattern Recognition Letters*, vol. 23, pp. 1711-1719, 2002
- [27] Yilmaz, A., Gokmen, M. , "Eigenhill vs. eigenface and eighenedge", *Pattern Recognition*, vol. 34, pp. 181-184, 2001
- [28] Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J., "Face Recognition: A Literature Survey", *ACM Computing Surveys*, pp. 399-458, 2003