

METODE INTELIGENTE DE REZOLVARE A PROBLEMELOR REALE



Laura Dioşan
Tema 6

Calcul afectiv (*Affective Computing*)

■ Scop

- Detectarea și recunoașterea emoțiilor
- Construirea mașinilor emoționale

■ Problematika științifică

- Vorbire emotivă → analiză și recunoaștere de semnal vocal
- Mimică emotivă → analiză și recunoaștere de informații vizuale
- Gesturi emotive → analiza și recunoașterea gesturilor
- Monitorizare psihologică

■ Domenii de aplicare

- E-learning
- Robotică
- Dispozitive personalizate (Siri, Kinect, Jocuri)

Calcul afectiv (*Affective Computing*)

□ Termenul *Affective computing*

- Introdus de Roz Picard în 1995 (a se vedea "*Affective Computing*", 1997)
- Definiție: "*computing that relates to, arises from, and deliberately influences emotion*"

□ Azi – o comunitate

- Societate profesională (Association for the Advancement of Affective Computing)
- Conferință internațională (ACII)
- Revistă (IEEE Transactions on Affective Computing)

□ Mai multe povești:

- <https://cs.uwaterloo.ca/~jhoey/teaching/cs886-affect/schedule.html>

Calcul afectiv (*Affective Computing*)

□ Recunoașterea emoțiilor în

■ Vorbire

- Emoții ale vorbirii naturale
- Detectarea depresiei

■ Texte

- Opinii enunțate pe bloguri (Twitter)
- Emoticoane

■ Mimica facială

- Înțelegerea impactului îmbătrânirii

■ Psihologie

- Interferarea muzică – activitate cerebrală (electro-encefalogramă)
- Detectarea stresului pe baza conductanței dermale
 - **Activitate electordermală (EDA)** – proprietate a corpului uman de a cauza variații continue în caracterizarea electrică a pielii (**galvanic skin response (GSR), electrodermal response (EDR)**)

■ Jocuri sau amuzamente computaționale

- Răspunsuri la câștiguri sau învingeri
- Affective music player
- Detectarea stărilor de plictiseală

■ Modelare

- Modelarea influenței emoțiilor asupra luării decizilor
- Modelarea factorilor care determină apariția emoțiilor
- Modelarea comportamentelor

Detectarea emoțiilor în vorbire

□ Ideea de bază

- Vorbirea este afectată de emoții
 - Teama sau furia – vorbire rapidă și puternic articulată, cu tonalitate ridicată
 - Oboseala, plictiseala sau supărarea – vorbire lentă și slab articulată, cu tonalitate redusă
- Procesarea emoțiilor din vorbire presupune
 - recunoașterea emoțiilor vorbitorului prin analiza unor șabloane ale vorbirii
 - analiza unor parametri vocali și caracteristici prozodice

□ Sistem de recunoaștere a emoțiilor în vorbire - etape

- Extragerea atributelor din datele (audio)
- Clasificarea emoțiilor emise în timpul vorbirii

Detectarea emoțiilor în vorbire

□ Ideea de bază

- Vorbirea este afectată de emoții
 - Teamă sau furie – vorbire rapidă și puternic articulată, cu tonalitate ridicată
 - Oboseală, plictiseală sau supărare – vorbire lentă și slab articulată, cu tonalitate redusă
- Procesarea emoțiilor din vorbire presupune
 - recunoașterea emoțiilor vorbitorului prin analiza unor șabloane ale vorbirii
 - analiza unor parametri vocali și caracteristici prozodice

□ Sistem de recunoaștere a emoțiilor în vorbire - etape

- **Extragerea atributelor din datele (audio)**
- Clasificarea emoțiilor emise în timpul vorbirii

Detectarea emoțiilor în vorbire

□ Baze de date

- Evaluare – cât de bine se simulează situațiile reale
 - Emoții reale sau regizate?
 - Cine (actori profesioniști sau semi-profesioniști) emite emoțiile?
 - Cum se simulează emoțiile?
 - Vorbitori experimentați (actori)
 - Wizard-of-Oz – interacțiune cu un computer (jocuri)
 - Emiterea emoțiilor este echilibrată sau nu?
 - Emiterea este uniform distribuită peste toate emoțiile?
 - Se rostesc aceleași lucruri, dar cu emoții diferite?

Detectarea emoțiilor în vorbire

- Baze de date – M. Ayadi, M. S. Kamel, F. Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition 44 (2011) 572–587

Corpus	Access	Language	Size	Source	Emotions
LDC Emotional Prosody Speech and Transcripts [78]	Commercially available	English	7 actors × 15 emotions × 10 utterances	Professional actors	Neutral, panic, anxiety, hot anger, cold anger, despair, sadness, elation, joy, interest, boredom, shame, pride, contempt
Berlin emotional database [18]	Public and free	German	800 utterances (10 actors × 7 emotions × 10 utterances + some second version) = 800 utterances	Professional actors	Anger, joy, sadness, fear, disgust, boredom, neutral
Danish emotional database [38]	Public with license fee	Danish	4 actors × 5 emotions (2 words + 9 sentences + 2 passages)	Nonprofessional actors	Anger, joy, sadness, surprise, neutral
Natural [91]	Private	Mandarin	388 utterances, 11 speakers, 2 emotions	Call centers	Anger, neutral
ESMBS [94]	Private	Mandarin	720 utterances, 12 speakers, 6 emotions	Nonprofessional actors	Anger, joy, sadness, disgust, fear, surprise
INTERFACE [54]	Commercially available	English, Slovenian, Spanish, French	English (186 utterances), Slovenian (190 utterances), Spanish (184 utterances), French (175 utterances)	Actors	Anger, disgust, fear, joy, surprise, sadness, slow neutral, fast neutral
KISMET [15]	Private	American English	1002 utterances, 3 female speakers, 5 emotions	Nonprofessional actors	Approval, attention, prohibition, soothing, neutral
BabyEars [120]	Private	English	509 utterances, 12 actors (6 males + 6 females), 3 emotions	Mothers and fathers	Approval, attention, prohibition
SUSAS [140]	Public with license fee	English	16,000 utterances, 32 actors (13 females + 19 males)	Speech under simulated and actual stress	Four stress styles: Simulated Stress, Calibrated Workload Tracking Task, Acquisition and Compensatory Tracking Task, Amusement Park Roller-Coaster, Helicopter Cockpit Recordings
MPEG-4 [114]	Private	English	2440 utterances, 35 speakers	U.S. American movies	Joy, anger, disgust, fear, sadness, surprise, neutral
Beihang University [43]	Private	Mandarin	7 actors × 5 emotions × 20 utterances	Nonprofessional actors	Anger, joy, sadness, disgust, surprise
FERMUS III [112]	Public with license fee	German, English	2829 utterances, 7 emotions, 13 actors	Automotive environment	Anger, disgust, joy, neutral, sadness, surprise
KES [65]	Private	Korean	5400 utterances, 10 actors	Nonprofessional actors	Neutral, joy, sadness, anger
CLDC [146]	Private	Chinese	1200 utterances, 4 actors	Nonprofessional actors	Joy, anger, surprise, fear, neutral, sadness
Hao Hu et al. [56]	Private	Chinese	8 actors × 5 emotions × 40 utterances	Nonprofessional actors	Anger, fear, joy, sadness, neutral
Amir et al. [2]	Private	Hebrew	60 Hebrew and 1 Russian actors	Nonprofessional actors	Anger, disgust, fear, joy, neutral, sadness
Pereira [55]	Private	English	2 actors × 5 emotions × 8 utterances	Nonprofessional actors	Hot anger, cold anger, joy, neutral, sadness

Detectarea emoțiilor în vorbire

- Extragerea atributelor din datele (audio)
 - Atribute locale
 - semnalul audio este împărțit în mai multe segmente mici (frames)
 - în cadrul unui segment, semnalul este considerat staționar
 - Se calculează atribute locale pt fiecare segment
 - atributele prozodice (ritm, anergie)
 - Atribute globale
 - Statistici calculate pentru o întreagă rostire
 - Abordare mixtă – un sunet (ex. o vocală) poate fi pronunțat diferit, în funcție de emoțiile trăite
 - semnalul audio se împarte în mai multe segmente bazate pe foneme
 - se calculează un singur vector de atribute pentru fiecare fenomen segmentat
 - Atribute locale vs. atribute globale
 - Avantaje pt atributele globale
 - Acuratețea clasificării
 - mai ales pentru emoțiile puternice-excitatoare (furie, teamă, bucurie) versus emoțiile slab-excitatoare (supărare)
 - Timpul necesar clasificării
 - Număr redus
 - Avantaje pt atributele locale
 - Nu lucrează bine pentru emoțiile cu același grad de excitație (furie versus bucurie)
 - Nu se pierde informația temporală (ca la atributele globale)

Detectarea emoțiilor în vorbire

- ❑ Extragerea atributelor din datele (audio)
 - Atribute de frecvență
 - ❑ Forma accentului – afectată de ritmul schimbării frecvenței
 - ❑ Tonul mediu – cât de tare se vorbește (relativ la vorbirea normală)
 - ❑ Pana conturului – tendința frecvenței de a se schimba de-a lungul timpului (crescătoare, descrescătoare, constantă)
 - ❑ Coborârea finală – descreșterea finală a frecvenței la finalul unei rostiri
 - ❑ Domeniul tonului – distanța între frecvența minimă și frecvența maximă a unei rostiri (a unui semnal)
 - Atribute temporale
 - ❑ Rata vorbirii – numărul de cuvinte/silabe rostite într-un interval de timp
 - ❑ Frecvența stresului – rata de apariție a tonurilor accentuate
 - Atribute calitative (parametrii pentru calitatea vocii și descrierea energiei)
 - ❑ Murmurul vocii (Breathiness) – măsoară zgomotul de aspirație în vorbire
 - ❑ Strălucirea (brilliance) – măsoară dominanța frecvențelor joase sau înalte
 - ❑ Intensitatea (loudness) – măsoară amplitudinea semnalului audio (energia emisiei vocale)
 - ❑ Discontinuitatea pauzei (pause discontinuity) – descrie tranziția între sunete și liniște
 - ❑ Discontinuitatea tonului (pitch discontinuity) – descrie tranziția între frecvențe

Detectarea emoțiilor în vorbire

- Extragerea atributelor din datele (audio)
 - Attribute continue (dependente de timp)
 - Frecvențe fundamentale
 - Media
 - Mediana
 - Deviația standard
 - Frecvența maximă/minimă
 - Întinderea (min-max)
 - Coeficientți de regresie liniară
 - Parametri Legendre de ordin 4
 - Vibrațiile
 - Media diferențelor de ordin 1
 - Media modulelor diferențelor de ordin 1
 - Bruiajul
 - Raportul dintre numărul de tonuri crescătoare și numărul de tonuri descrescătoare
 - Energia
 - Media
 - Mediana
 - Deviația standard
 - Energia minimă/maximă
 - Întinderea (min-max)
 - Coeficientți de regresie liniară
 - Sclipirea (shimmer)
 - Parametri Legendre de ordin 4
 - Durata
 - Rata vorbirii
 - Raportul dintre durata vorbirii și durata pauzelor
 - Cea mai lungă durată a vorbirii
 - Formanți (concentrare a energiei acustice în jurul anumitor frecvențe)
 - Primul și al doilea formant și lățimea lor de bandă

Detectarea emoțiilor în vorbire

- Extragerea atributelor din datele (audio)
 - Calitative (ex. Aspră (Harsh), Tensionată (Tense), Breathy)
 - Nivelul vocii
 - Amplitudinea semnalului
 - Energia
 - Durata
 - Tonul vocii
 - Fraza, fenomenul, cuvintele și modalitățile de delimitare a cuvintelor/frazelor
 - Structurile temporale
 - Spectrale (reprezentare temporală a semnalului audio, dar pe termen-scurt)
 - ordinary linear predictor coefficients (LPC)
 - Mel-frequency cepstrum coefficients (MFCC)
 - Log-frequency power coefficients (LFPC)
 - TEO-based
 - TEO-decomposed FM variation (TEO-FM-Var)
 - normalized TEO autocorrelation envelope area (TEO-Auto-Env)
 - critical band-based TEO autocorrelation envelope area (TEO-CB-Auto-Env)

Detectarea emoțiilor în vorbire

□ Ideea de bază

- Vorbirea este afectată de emoții
 - Teama sau furia – vorbire rapidă și puternic articulată, cu tonalitate ridicată
 - Oboseala, plictiseala sau supărarea – vorbire lentă și slab articulată, cu tonalitate redusă
- Procesarea emoțiilor din vorbire presupune
 - recunoașterea emoțiilor vorbitorului prin analiza unor șabloane ale vorbirii
 - analiza unor parametri vocali și caracteristici prozodice

□ Sistem de recunoaștere a emoțiilor în vorbire - etape

- Extragerea atributelor din datele (audio)
- **Clasificarea emoțiilor emise în timpul vorbirii**

Detectarea emoțiilor în vorbire

- Clasificarea emoțiilor emise în timpul vorbirii – Algoritmi
 - linear discriminant classifiers LDC
 - Classification happens based on the value obtained from the linear combination of the feature values, which are usually provided in the form of vector features.
 - k-NN
 - Classification happens by locating the object in the feature space, and comparing it with the k nearest neighbours (training examples). The majority vote decides on the classification.
 - Gaussian mixture model GMM
 - is a probabilistic model used for representing the existence of sub-populations within the overall population. Each sub-population is described using the mixture distribution, which allows for classification of observations into the sub-populations.^[17]
 - SVM
 - is a type of (usually binary) linear classifier which decides in which of the two (or more) possible classes, each input may fall into.
 - ANN
 - is a mathematical model, inspired by biological neural networks, that can better grasp possible non-linearities of the feature space.
 - Decision tree algorithms
 - work based on following a decision tree in which leaves represent the classification outcome, and branches represent the conjunction of subsequent features that lead to the classification.
 - HMMs
 - a statistical Markov model in which the states and state transitions are not directly available to observation. Instead, the series of outputs dependent on the states are visible. In the case of affect recognition, the outputs represent the sequence of speech feature vectors, which allow the deduction of states' sequences through which the model progressed. The states can consist of various intermediate steps in the expression of an emotion, and each of them has a probability distribution over the possible output vectors. The states' sequences allow us to predict the affective state which we are trying to classify, and this is one of the most commonly used techniques within the area of speech affect detection.

De ce modele Markov?

- Modele de predicție pe baza unor observații (trecute)
 - Observații independente/corelate
 - Observații identic distribuite / cu distribuții diferite

- Ex.
 - Predicția vremii
 - Predicția cursurilor valutare
 - Recunoașterea de semnal vocal
 - Compoziția structurilor ADN

De ce modele Markov?

□ Modele Markov

■ Intrări

□ Model:



□ Observații:



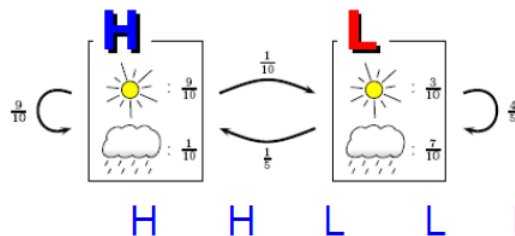
■ Ieșiri

□ $P(\text{Sunny Sunny Rainy Sunny Sunny}) = ?$

□ Modele Markov ascunse

■ Intrări

□ Model:



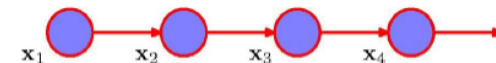
□ Observații:



■ Ieșiri

□ $P(\text{Sunny Sunny Rainy Sunny Sunny}) = ?$

Model Markov



□ Ideea de bază

- Similar unui automat finit, dar cu tranziție probabilistică de la o stare la alta
- La fiecare moment de timp se realizează o tranziție dintr-o stare în altă stare
 - La orice moment de timp, un eveniment se poate afla într-o singură stare
 - Fiecare eveniment are o singură ieșire (un singur rezultat)

□ Elemente

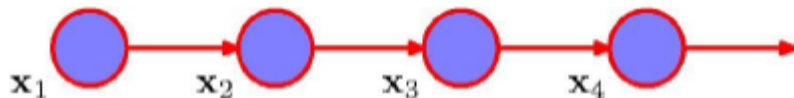
- Ceas \rightarrow cuantifică momentele de timp (discret)
 - $t = \{1, 2, 3, \dots, T\}$
- N stări
 - $Q = \{S_1, S_2, \dots, S_N\}$
- N evenimente
 - $E = \{e_1, e_2, \dots, e_N\}$
- Probabilități inițiale
 - $\Pi_j = P(S_1 = S_j), 1 \leq j \leq N$ $\sum_{j=1}^N \pi_j = 1$
- Probabilități de tranziție dintr-o stare S_i (de la momentul $t-1$) într-o stare S_j (de la momentul t)
 - $a_{ij} = P[q_t = S_j \mid q_{t-1} = S_i] \quad 1 \leq i, j \leq N$
 - Obs.: suma probabilităților care ies dintr-o stare = 1

Model Markov

□ Cazuri particulare

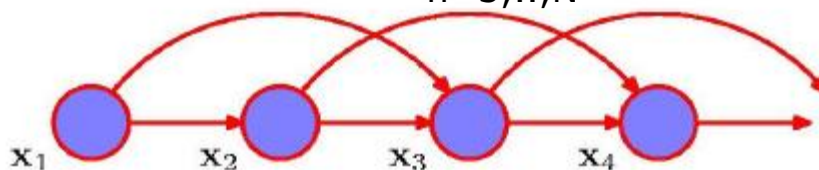
■ MM de ordin 1

- A n -a observație este influențată doar de a $n-1$ -a observație
- $p(x_n|x_1, x_2, \dots, x_{n-1}) = p(x_n|x_{n-1})$
- $p(x_1, x_2, \dots, x_N) = p(x_1) \prod_{n=2, \dots, N} p(x_n|x_{n-1})$



■ MM de ordin 2

- A n -a observație este influențată de a $n-1$ -a observație și de a $n-2$ -a observație
- $p(x_n|x_1, x_2, \dots, x_{n-1}) = p(x_n|x_{n-1}, x_{n-2})$
- $p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2|x_1) \prod_{n=3, \dots, N} p(x_n|x_{n-1}, x_{n-2})$



Model Markov

□ Exemple

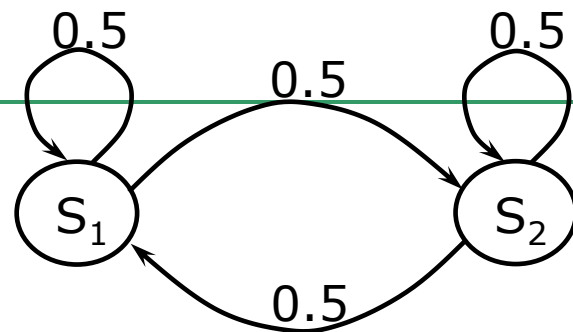
■ Aruncarea unei monezi perfecte

□ Intrări:

- Starea $S_1 \rightarrow$ evenimentul "moneda cade și se vede capul (C)"
- Starea $S_2 \rightarrow$ evenimentul "moneda cade și se vede pajura (P)"
- Probabilitatea apariției:
 - Stării S_1 : $P(C) = \frac{1}{2} = 0.5$
 - Stării S_2 : $P(P) = \frac{1}{2} = 0.5$
 - Stării S_1 după starea S_1 : $P(C|C) = \frac{1}{2} = 0.5$
 - Stării S_1 după starea S_2 : $P(C|P) = \frac{1}{2} = 0.5$
 - Stării S_2 după starea S_2 : $P(P|P) = \frac{1}{2} = 0.5$
 - Stării S_2 după starea S_1 : $P(P|C) = \frac{1}{2} = 0.5$

□ Ieșiri:

- Care este probabilitatea generării unei secvențe de evenimente C P C C P C P P C C?
 - A unei secvențe de sări $S_1 S_2 S_1 S_1 S_2 S_1 S_2 S_2 S_2 S_1 S_1$



Model Markov

□ Exemple

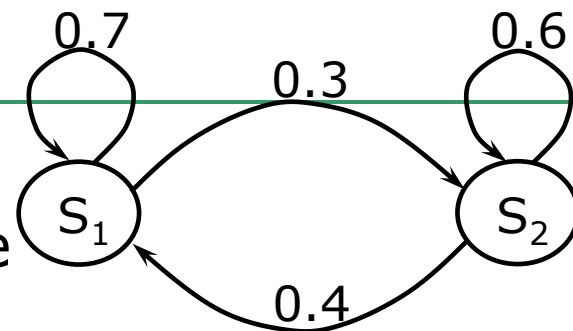
■ Aruncarea unei monezi imperfecte

□ Intrări:

- Starea $S_1 \rightarrow$ evenimentul "moneda cade și se vede capul (C)"
- Starea $S_2 \rightarrow$ evenimentul "moneda cade și se vede pajura (P)"
- Probabilitatea apariției:
 - Stării S_1 : $P(C) = \frac{1}{2} = 0.5$
 - Stării S_2 : $P(P) = \frac{1}{2} = 0.5$
 - Stării S_1 după starea S_1 : $P(C|C) = 0.7$
 - Stării S_1 după starea S_2 : $P(C|P) = 0.4$
 - Stării S_2 după starea S_2 : $P(P|P) = 0.6$
 - Stării S_2 după starea S_1 : $P(P|C) = 0.3$

□ Ieșiri:

- Care este probabilitatea generării unei secvențe de evenimente C P C C P C P P C C?
 - A unei secvențe de sări $S_1 S_2 S_1 S_1 S_2 S_1 S_2 S_2 S_2 S_1 S_1$



Model Markov

Exemple

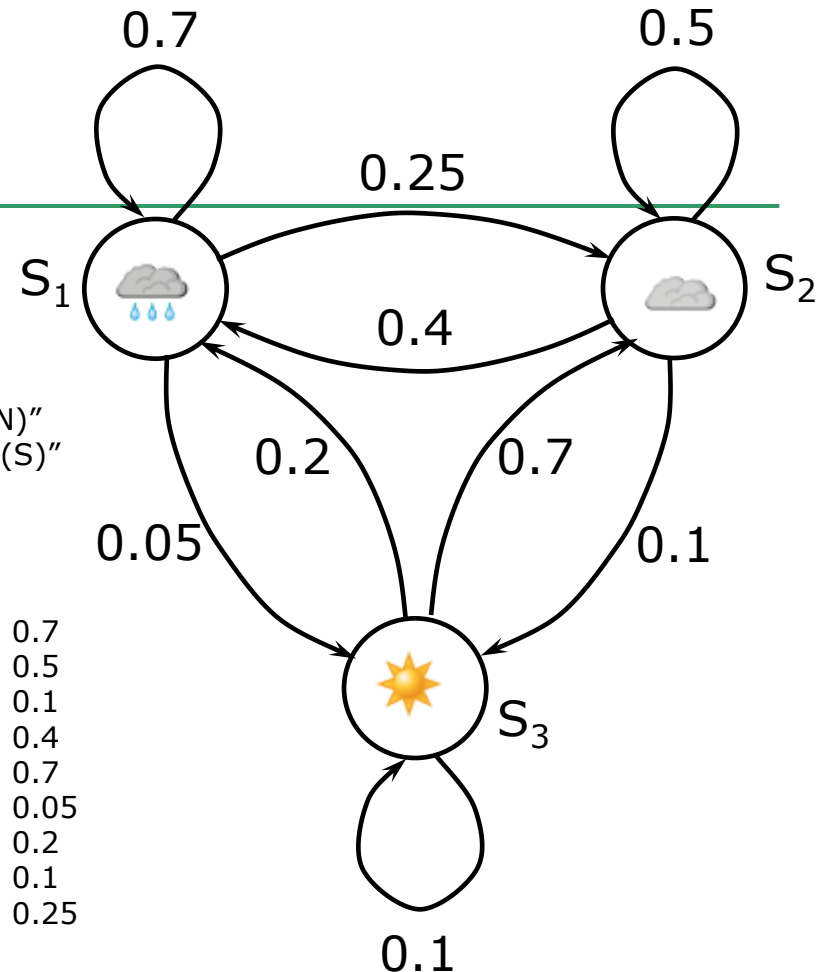
Previziunea vremii

Intrări:

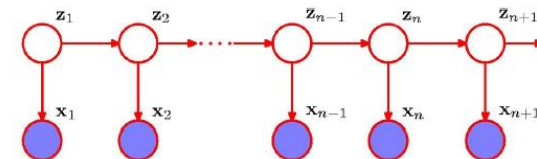
- Starea S1 → evenimentul "afară plouă (P)"
- Starea S2 → evenimentul "afară sunt nori (N)"
- Starea S3 → evenimentul "afară este soare (S)"
- Probabilitatea apariției:
 - Stării S1: $P(S1) = 1/3 = 0.5$
 - Stării S2: $P(S2) = 1/3 = 0.4$
 - Stării S3: $P(S3) = 1/3 = 0.1$
 - Stării S1 după starea S1: $P(S1|S1) = 0.7$
 - Stării S2 după starea S2: $P(S2|S2) = 0.5$
 - Stării S3 după starea S3: $P(S3|S3) = 0.1$
 - Stării S1 după starea S2: $P(S1|S2) = 0.4$
 - Stării S2 după starea S3: $P(S2|S3) = 0.7$
 - Stării S3 după starea S1: $P(S3|S1) = 0.05$
 - Stării S1 după starea S3: $P(S1|S3) = 0.2$
 - Stării S3 după starea S2: $P(S3|S2) = 0.1$
 - Stării S2 după starea S1: $P(S2|S1) = 0.25$

Ieșiri:

- Care este probabilitatea generării unei secvențe de evenimente P P P N S N P?
 - A unei secvențe de sări S3 S3 S3 S2 S1 S2 S3?
 - $P = P[S_1] P[S_1|S_1] P[S_1|S_1] P[S_2|S_1] P[S_3|S_2] P[S_2|S_3] P[S_1|S_2]$
 - $P = 0.5 \cdot 0.7 \cdot 0.7 \cdot 0.25 \cdot 0.1 \cdot 0.7 \cdot 0.4$
 - $P = 0.001715$
- Care este probabilitatea generării unei secvențe de evenimente S S S P N S S?
 - A unei secvențe de sări S3, S3, S3, S1, S2, S3, S3?
 - $P = P[S_3] P[S_3|S_3] P[S_3|S_3] P[S_1|S_3] P[S_2|S_1] P[S_3|S_2] P[S_3|S_3]$
 - $P = 0.1 \cdot 0.1 \cdot 0.1 \cdot 0.2 \cdot 0.25 \cdot 0.1 \cdot 0.1$
 - $P = 5.0 \times 10^{-7}$



Model Markov ascuns



□ Ideea de bază

- Similar unui automat finit, dar cu tranziție probabilistică de la o stare la alta
- La fiecare moment de timp se realizează o tranziție dintr-o stare în altă stare
 - La orice moment de timp, un eveniment se poate afla în mai multe stări (cu o anumită probabilitate)

□ Elemente

- Ceas → cuantifică momentele de timp (discret)
 - $t = \{1, 2, 3, \dots, T\}$
- N stări
 - $Q = \{S1, S2, \dots, SN\}$
- M evenimente
 - $E = \{e1, e2, \dots, eM\}$
- Probabilități inițiale
 - $\Pi_j = P(S1=Sj), 1 \leq j \leq N$ $\sum_{j=1}^N \pi_j = 1$
- Probabilități de tranziție dintr-o stare S_i (de la momentul $t-1$) într-o stare S_j (de la momentul t)
 - $a_{ij} = P[q_t = S_j \mid q_{t-1} = S_i] \quad 1 \leq i, j \leq N$
 - Obs.: suma probabilităților care iese dintr-o stare = 1
- Probabilități ale observațiilor
 - $b_j(k) = P[\mathbf{o}_t = e_k \mid q_t = S_j] \quad 1 \leq k \leq M$
 - $b_j(\mathbf{o}_t) = P[\mathbf{o}_t = e_k \mid q_t = S_j] \quad 1 \leq k \leq M$

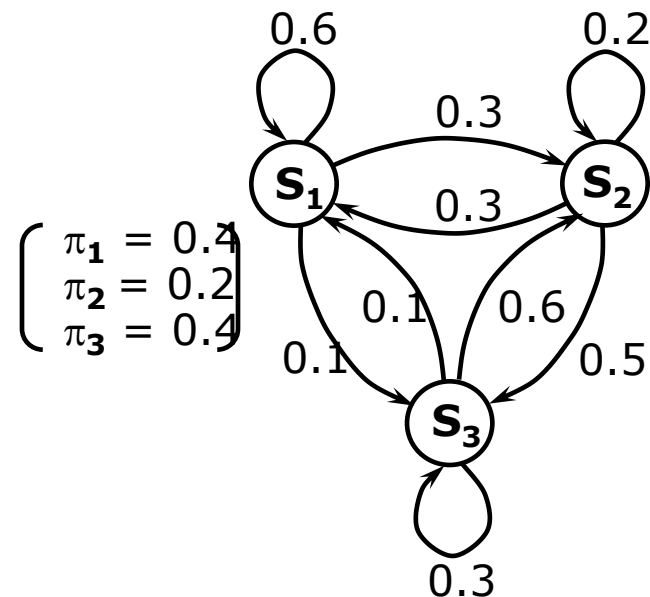
□ Model Markov ascuns = (π, A, B) , unde

- $\Pi \rightarrow$ probabilitățile stărilor inițiale
- A – matricea probabilităților de tranziție de la o stare ascunsă la alta ($N \times N$)
- B – matricea probabilităților de tranziție de la o stare ascunsă la o observație ($N \times M$)

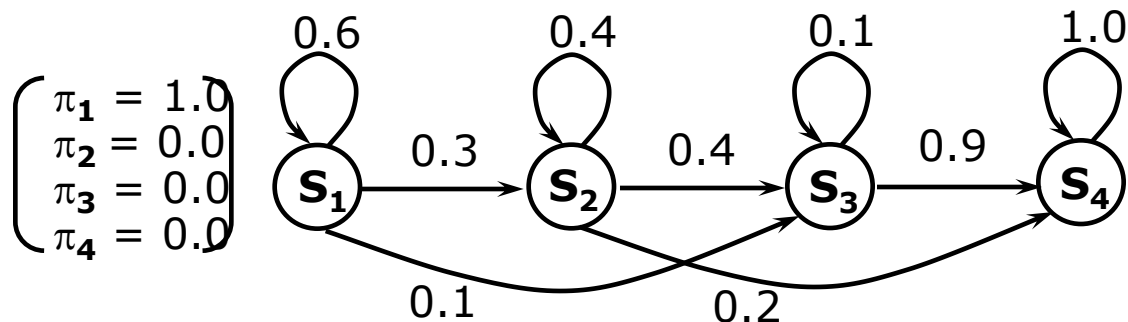
Model Markov ascuns

□ Cazuri particulare

■ Ergodice (complet conectate)



■ Bakis (stânga-dreapta)



Model Markov ascuns

Exemple

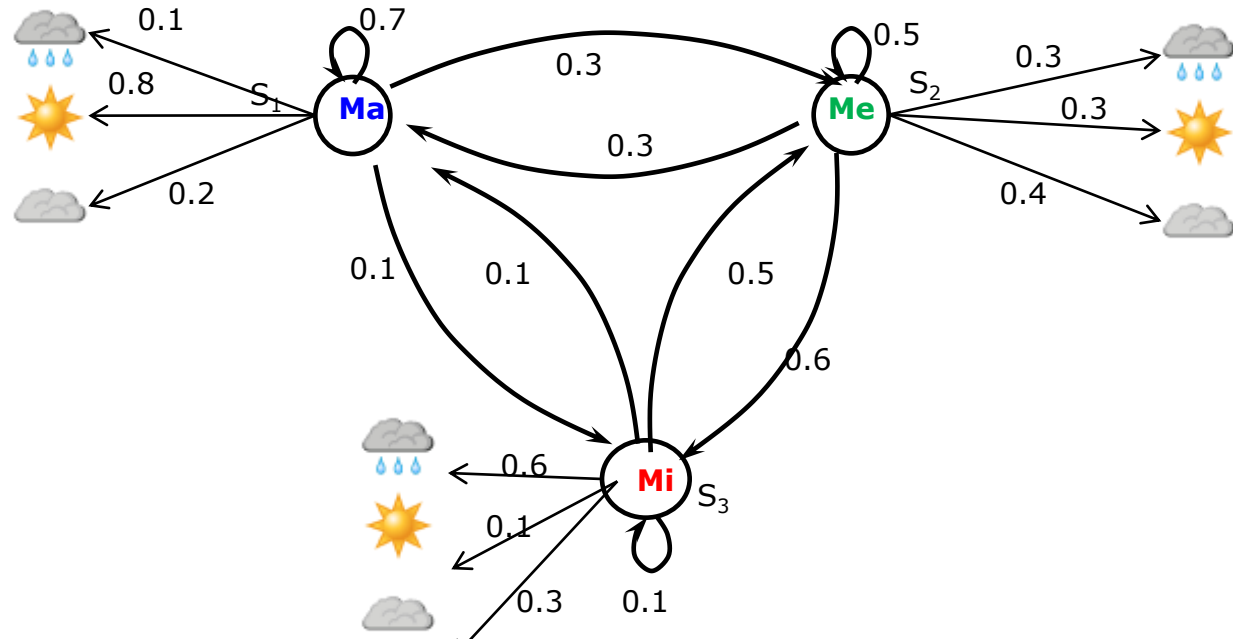
Previțiunea vremii

Intrări:

- Starea S1 → evenimentul "afară plouă (P)"
- Starea S2 → evenimentul "afară sunt nori (N)"
- Starea S3 → evenimentul "afară este soare (S)"
- Probabilitatea apariției unei
 - presiuni atmosferice mare: $\pi_{Ma} = 0.4$
 - presiuni atmosferice medie: $\pi_{Me} = 0.2$
 - presiuni atmosferice mică: $\pi_{Mi} = 0.3$
 - presiuni atmosferice mare după presiune atmosferică medie: $P(Ma|Me) = a_{Me,Ma} = 0.3$
 - presiuni atmosferice mare după presiune atmosferică mică: $P(Ma|Mi) = a_{Mi,Ma} = 0.1$
 - presiuni atmosferice medie după presiune atmosferică mare: $P(Me|Ma) = a_{Ma,Me} = 0.3$
 - presiuni atmosferice medie după presiune atmosferică mică: $P(Me|Mi) = a_{Mi,Me} = 0.6$
 - presiuni atmosferice mică după presiune atmosferică mare: $P(Mi|Ma) = a_{Ma,Mi} = 0.1$
 - presiuni atmosferice mică după presiune atmosferică medie: $P(Mi|Me) = a_{Me,Mi} = 0.5$
 - presiuni atmosferice mare după presiune atmosferică mare: $P(Ma|Ma) = a_{Ma,Ma} = 0.6$
 - presiuni atmosferice medie după presiune atmosferică medie: $P(Me|Me) = a_{Me,Me} = 0.2$
 - presiuni atmosferice mică după presiune atmosferică mică: $P(Mi|Mi) = a_{Mi,Mi} = 0.3$

π	
Ma	0.4
Me	0.2
Mi	0.3

a	Ma	Me	Mi
Ma	0.6	0.3	0.1
Me	0.3	0.2	0.5
Mi	0.1	0.6	0.3



Model Markov ascuns

Exemple

Previziziunea vremii

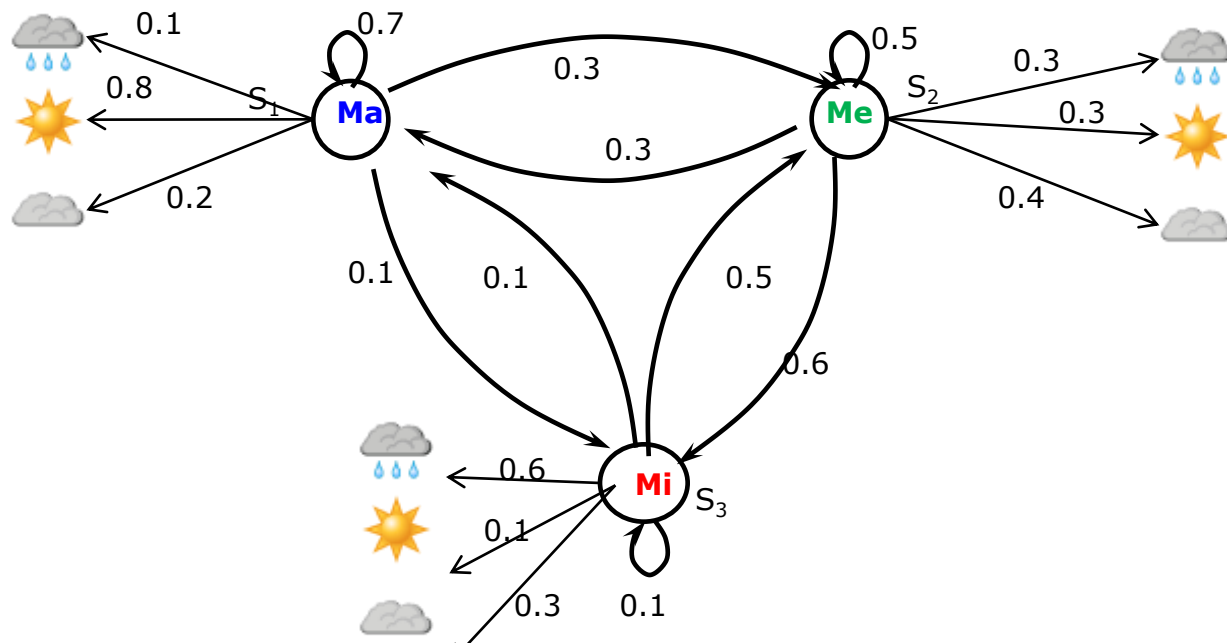
Intrări:

- Starea S1 → evenimentul "afară plouă (P)"
- Starea S2 → evenimentul "afară sunt nori (N)"
- Starea S3 → evenimentul "afară este soare (S)"
- Probabilitatea apariției:
 - Ploii și a unei presiuni atmosferice mare $P(P|Ma)=b_{Ma,P}=0.1$
 - Ploii și a unei presiuni atmosferice medie $P(P|Me)=b_{Me,P}=0.3$
 - Ploii și a unei presiuni atmosferice mică $P(P|Mi)=b_{Mi,P}=0.6$
 - Norilor și a unei presiuni atmosferice mare $P(N|Ma)=b_{Ma,N}=0.2$
 - Norilor și a unei presiuni atmosferice medie $P(N|Me)=b_{Me,N}=0.4$
 - Norilor și a unei presiuni atmosferice mică $P(N|Mi)=b_{Mi,N}=0.3$
 - Soarelui și a unei presiuni atmosferice mare $P(S|Ma)=b_{Ma,S}=0.8$
 - Soarelui și a unei presiuni atmosferice medie $P(S|Me)=b_{Me,S}=0.3$
 - Soarelui și a unei presiuni atmosferice mică $P(S|Mi)=b_{Mi,S}=0.1$

π	
Ma	0.4
Me	0.2
Mi	0.3

a	Ma	Me	Mi
Ma	0.6	0.3	0.1
Me	0.3	0.2	0.5
Mi	0.1	0.6	0.3

b	S	N	P
Ma	0.8	0.2	0.1
Me	0.3	0.4	0.3
Mi	0.1	0.3	0.6



Model Markov ascuns

Exemple

Previziunea vremii

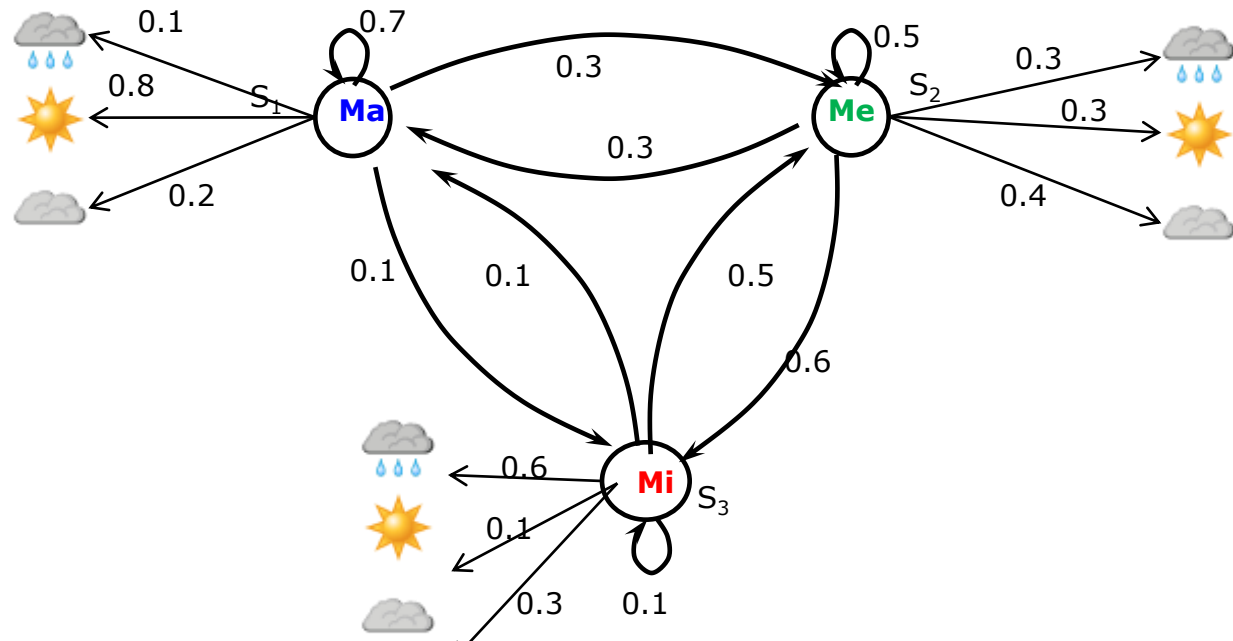
Ieșiri:

π	
Ma	0.4
Me	0.2
Mi	0.3

	Ma	Me	Mi
Ma	0.6	0.3	0.1
Me	0.3	0.2	0.5
Mi	0.1	0.6	0.3

b	S	N	P
Ma	0.8	0.2	0.1
Me	0.3	0.4	0.3
Mi	0.1	0.3	0.6

- Care este probabilitatea generării unei secvențe de evenimente P P P N S N P fiind date modelul HMM (π, A, B) și o secvență de stări ascunse (seq=MaMeMeMiMi Ma Ma)?
 - $P = P[P|Ma] P[P|Me] P[P|Me] P[N|Mi] P[S|Mi] P[N|Ma] P[P|Ma]$
 - $P = b_{Ma,P} b_{Me,P} b_{Me,P} b_{Mi,N} b_{Mi,S} b_{Ma,N} b_{Ma,P}$
 - $P = 0.1 * 0.3 * 0.3 * 0.3 * 0.1 * 0.2 * 0.1$
- Care este probabilitatea generării unei secvențe de evenimente P P P N S N P și a unei secvențe de stări ascunse (seq=Ma Me Me Mi Mi Ma Ma) fiind dat modelul HMM (π, A, B)?
 - $P = P_{start}[Ma] P[P|Ma] * P[Me|Ma] P[P|Me] * P[Me|Me] P[P|Me] * P[Mi|Me] P[N|Mi] * P[Mi|Mi] P[S|Mi] * P[Ma|Mi] P[N|Ma] * P[Ma|Ma] P[P|Ma]$
 - $P = \pi_{Ma} b_{Ma,P} a_{Ma,Me} b_{Me,P} a_{Me,Me} b_{Me,P} a_{Me,Mi} b_{Mi,N} a_{Mi,Mi} b_{Mi,S} a_{Mi,Me} b_{Me,N} a_{Ma,Me} b_{Ma,P}$
 - $P = 0.4 * 0.1 * 0.3 * 0.3 * 0.2 * 0.3 * 0.5 * 0.3 * 0.3 * 0.1 * 0.1 * 0.2 * 0.6 * 0.1$



Model Markov ascuns

□ Dileme

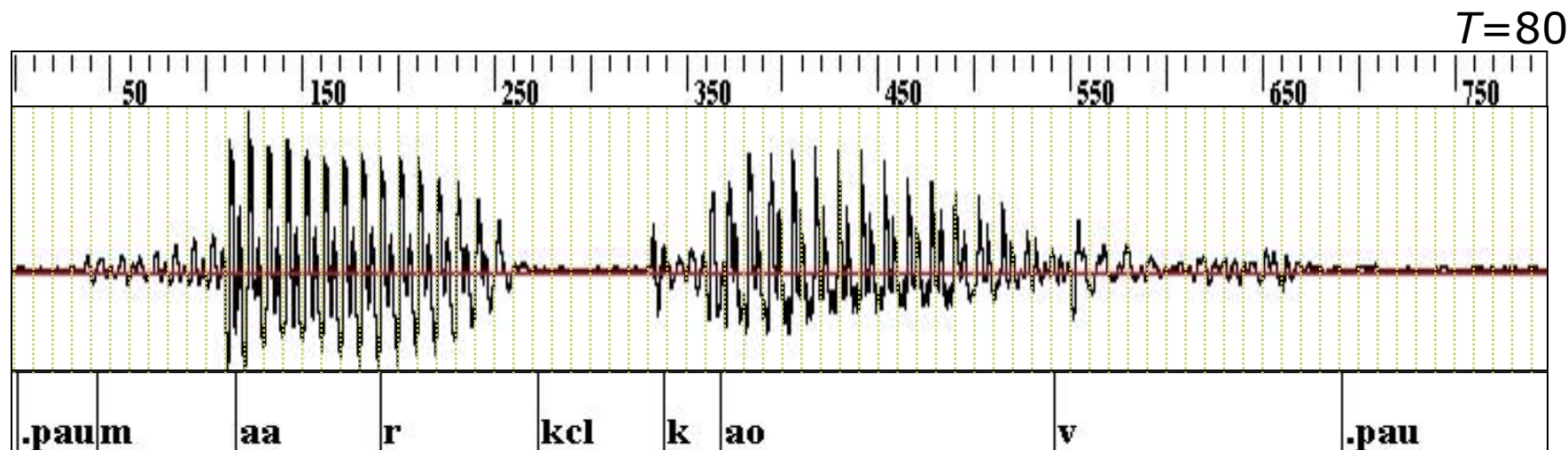
- Care este probabilitatea producerii unei secvențe de observații O?
 - Algoritmul forward-backward
 - Ex. care este probabilitatea ca săptămâna să fie S S N P S S S?
- Care este următoarea observație după o secvență de observații?
 - Calcul de probabilități condiționate
 - Ex. Care este cea mai probabilă vreme pentru mâine știind că ultima săptămână a fost S S N P S S S?
- Care este cea mai probabilă explicație (secvență de stări ascunse) care a generat un set de observații?
 - Algoritmul Viterbi
 - Ex. Care este cea mai probabilă secvență de nivele ale presiunii (Mică, Medie, Mare) care au generat o săptămână cu S S N P S S S?

Model Markov ascuns

□ Exemple

■ Recunoașterea vocii

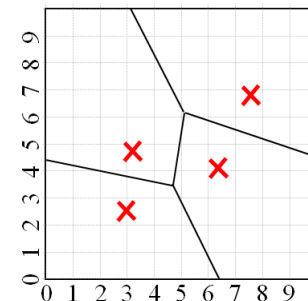
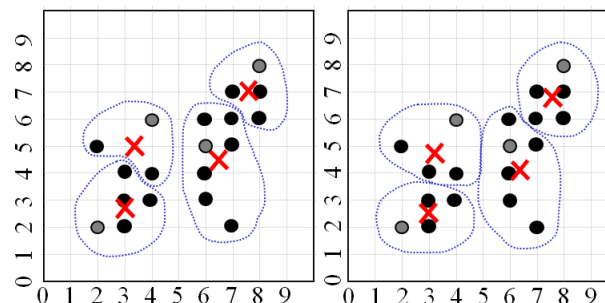
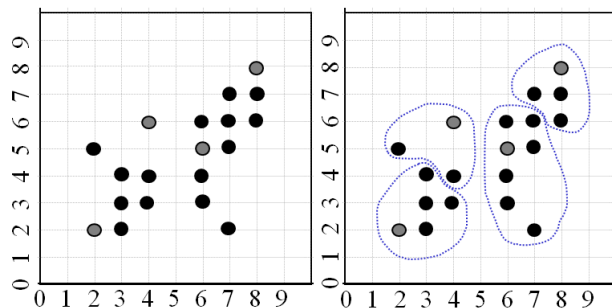
- Semnalul vocal este separat în fragmente de 10-msec
 - Cum? → Quantificare vectorială
- generându-se o mulțime O de observații (stări observabile).
- Pe baza HMM-urilor corespunzătoare fiecărui posibil cuvânt w_i (fonem f_i), se calculează probabilitatea P_i ca observațiile O să fi fost generate de cuvântul w_i (fonemul f_i).
- Se alege cea mai mare probabilitate → cuvântul rostit (fonemul rostit).



Cuantificare vectorială

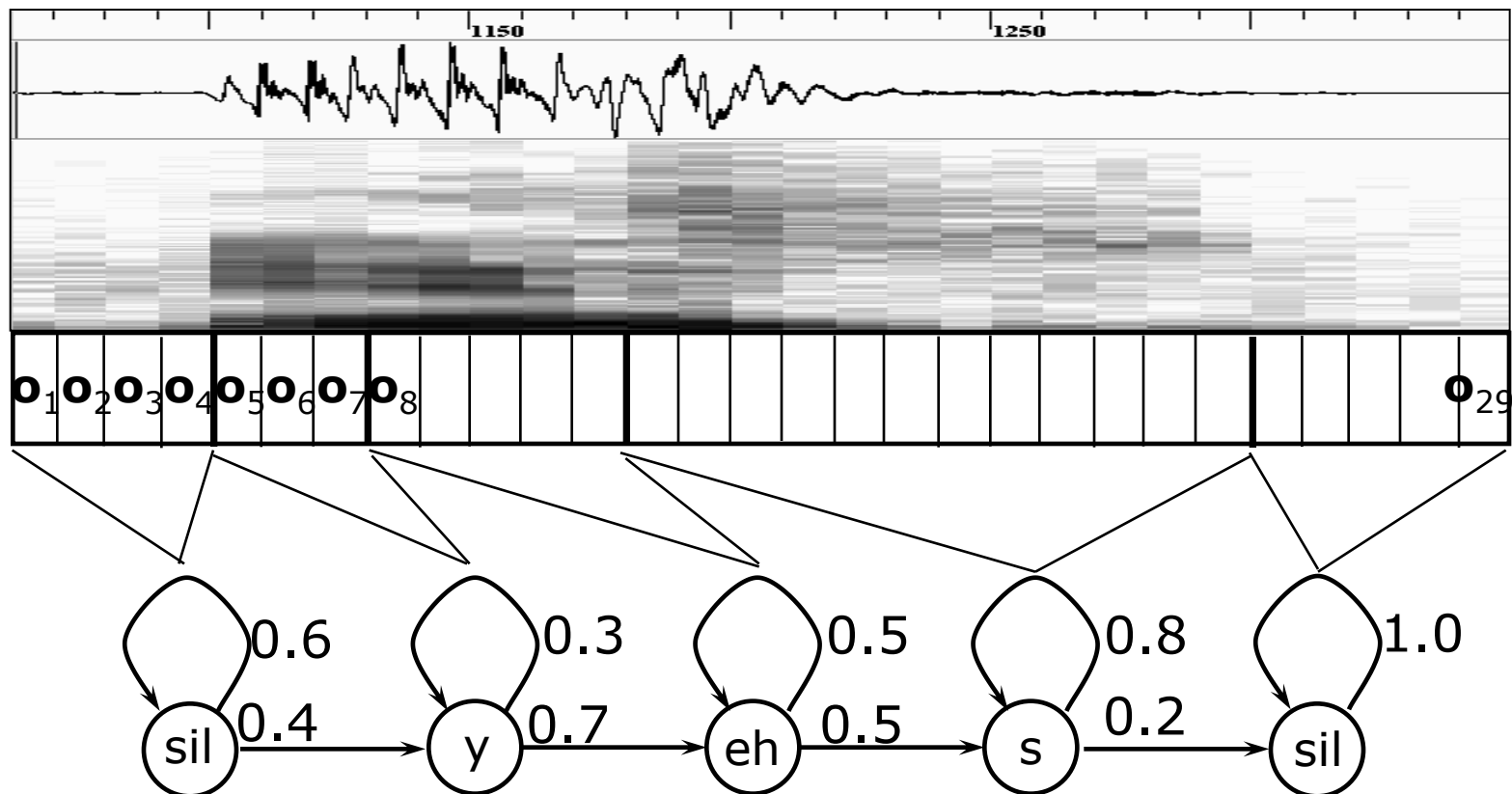
□ Ideea de bază

- Separarea spațiului de caracteristici în grupe de date
 - Fiecare grupă are asociat un centroid
 - Toți centroizii formează un *codebook*
- cu scopul reducerii dimensiunii datelor



Cuantificarea vectorială și generarea observațiilor

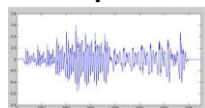
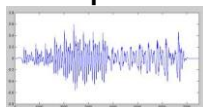
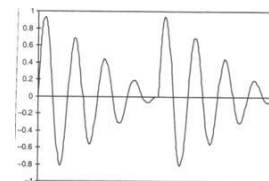
- Generarea observațiilor corespunzătoare unei stări
- Starea =
 - Cuvânt
 - Mono-fonem (mono-gram)
 - Fonem independent (b)
 - Ex. BRYAN → **B R AY AX N**
 - Bi-fonem (bi-gram)
 - Fonele contextuale (stânga → a-b, dreapta → b+c)
 - Ex. BRYAN → SIL-**B B-R R-AY AY-AX AX-N**, BRyAN → **B+R R+AY AY+AX AX+N N+SIL**
 - Tri-fonem (tri-gram)
 - Fonele contextuale (stânga și dreapta → a-b+c)
 - Ex. BRYAN → SIL-**B+R B-R+AY R-AY+AX AY-AX+N AX-N+SIL**



Model Markov ascuns

□ Dileme

- Care este probabilitatea producerii unei secvențe de observații O ?
 - Algoritmul forward-backward
 - Ex.
 - Care este probabilitatea ca semnalul vocal să fie ?
- Care este următoarea observație după o secvență de observații?
 - Calcul de probabilități condiționate
 - Ex.
 - Care este cel mai probabil semnal vocal următor știind că semnalul curent a fost ?
- Care este cea mai probabilă explicație (secvență de stări ascunse) care a generat un set de observații?
 - Algoritmul Viterbi
 - Ex.
 - Care este cea mai probabilă secvență de foneme care au generat un semnal audio ?



Detectarea emoțiilor în vorbire

□ Câteva instrumente

■ Praat

□ <http://www.fon.hum.uva.nl/praat/>

■ Speech Prosody Analysis Software Tools

□ <http://affect.media.mit.edu/software.php>

■ TKK Aparat

□ <https://sourceforge.net/projects/aparat/>

■ Virtual Human Toolkit

□ <https://vhtoolkit.ict.usc.edu/>

■ Litebody

□ <http://relationalagents.com/litebody.html>

Detectarea emoțiilor în gesturi

□ Gesturi

- Mimica feței
- Gesturi ale corpului

□ Emoțiile faciale

- Științele cognitive
 - Ce informație trebuie reprezentată?
- Științele computaționale
 - Cum trebuie reprezentată (codată) informația?

□ Percepția feței

- Abordarea holistică
 - Fața este modelată ca un întreg, fără părți (componente) care ar putea fi izolate
- Abordarea bazată pe componente
 - Anumite attribute ale feței pot fi procesate individual
- Abordarea bazată pe configurație
 - Se modelează relațiile spațiale dintre componentele feței (ex. ochiul stâng-ochiul drept, nas-gură)

Detectarea emoțiilor în expresia feței

- Recunoașterea afectivă are drept scop
 - Recunoașterea gesturilor de pe față
- Facial Action Coding System (FACS)
 - Compus din *facial Action Units* (AU) – coduri care descriu configurații ale feței (ex. AU 12 – colțul buzelor)
- Efectuarea unui gest are o evoluție temporală care se poate modela prin 4 segmente:
 - *Neutru* – faza fără expresii și activitate musculară
 - *Onset* – apar contracțiile musculare și cresc în intensitate
 - *Apex* – platou (intensitate constantă)
 - *Offset* – relaxarea musculaturii

Detectarea emoțiilor în expresia feței

- Recunoașterea afectivă are drept scop
 - Recunoașterea emoțiilor de pe față (consecințe ale gesturilor)
 - Emoții de bază
 - Modelul dezvoltat de Ekman&Co (1972): producerea și interpretarea anumitor expresii sunt adânc întipărite în creier și recunoaște universal (nu sunt elemente culturale, specifice unei nații)
 - Emoțiile produse de aceste expresii se pot modela în 6 clase:
 - Fericire
 - Tristețe
 - Surpriză
 - Teamă
 - Furie
 - Dezgust
 - Emoții derivate (1990)
 - Emoții pozitive & Emoții negative
 - Unele emoții nu sunt codate de mușchii feței
 - Amuzament
 - Dispreț
 - Mulțumire
 - Jenă
 - Excitare
 - Vinovăție
 - Mândrie
 - Alinare
 - Satisfacție
 - Plăcere senzorială
 - Rușine

Detectarea emoțiilor în expresia feței

- Provocări în recunoașterea emoțiilor pe fețe
 - Variații ale poziției capului
 - Variații ale iluminării
 - Erori de aliniere
 - Ocluzii
 - Deplasări de identitate

Detectarea emoțiilor în expresia feței

□ Competiții

■ Facial Expression Recognition (FERA)

□ <http://sspnet.eu/fera2015/>

□ <http://sspnet.eu/fera2017/>

■ Audio/Visual Emotion Challenges (AVEC)

□ <http://sspnet.eu/avec2016/>

Detectarea emoțiilor în expresia feței

□ Baze de date

■ Conținut

- Imagini singulare ale emoțiilor (de intensitate maximă)
- Secvențe de imagini & video corespunzătoare unei emoții

Detectarea emoțiilor în expresia feței

- ▣ Baze de date – detalii in Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro , Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 37, NO. 6, JUNE 2015

	Dataset	Access-ible	Application and Labels				Statistics and Properties				Baseline		
			BE	NBE	AU	DA	#Sub-jects	#Vid-eos	#Im-ages	frame-by-frame labels?	Res-ults	Regis-trati-on	Rep-resent.
Posed	CK [61]	Yes	6+N	-	✓(+T,+I[146] [†])	-	97	486	-	-	-	-	-
	GEMEP [7]	Yes	6+N	12	✓	-	10	7,000	-	-	✓	✓	✓
	ISL Frontal-View [146]	Yes	-	-	✓ +T	-	10	42	-	✓	-	-	-
	ISL Multi-View [145]	Yes	-	-	✓ +T	-	8	40	-	✓	-	✓	-
	Multi-PIE [47]	Not free	3+N	2	-	-	100	-	4,200	-	-	-	-
Posed & Non-posed	JAFFE [86]	Yes	6+N	-	-	-	10	-	213	-	-	-	-
	MMI [105], [153]	Yes	6+N	-	✓ +T	-	75	2,420	484	temp.phas.	-	-	-
	CK+ [83]	Yes	6+N	1	-	-	123	593	-	-	✓	✓	-
Non-posed	HUMAINE [91]	Yes	-	-	-	A/V*	4	23	-	✓	-	-	-
	SEMAINE [91]	Yes	3	10 ^{††}	✓	A/E/P/V*	150	959	-	✓	✓	✓	✓
	RU-FACS [10]	Partly	-	-	✓	-	100	100	-	N/A	-	-	-
	DISFA [10]	Yes	-	-	✓ +I	-	27	27	-	✓	✓	✓	-
	Belfast	Yes	6+N	Var ^{††}	-	A/V*	256	1,400	-	✓	-	-	-
	Induced [137]												
	Belfast	Yes	4+N	12	-	A/V*	125	298	-	✓	-	-	-
	Naturalistic [36]												
	GENKI-4K [143]	Yes	2	-	-	-	N/A	-	4,000	N/A	-	-	-
	UNBC-Mc Master [84]	Partly	-	Pain	✓ +I	-	25	200	-	✓	✓	✓	-
	COPE [15]	No	-	Pain	-	-	26	-	204	N/A	-	-	-
	SMIC [72]	Yes	3 [†] +N	✓	-	-	16	264	-	✓	✓	✓	-

[†]See text for details. ^{††}Refer to the original dataset paper for details. *These dimensions may be referred to with different names.

BE: Basic emotions; NBE: Non-basic emotions; AU: action units; DA: Dimensional affect;

N: Neutral; +T: Temporal segments; +I: AU intensity; A: Arousal; E: Expectancy; P: Power; V: Valence.

Detectarea emoțiilor în expresia feței

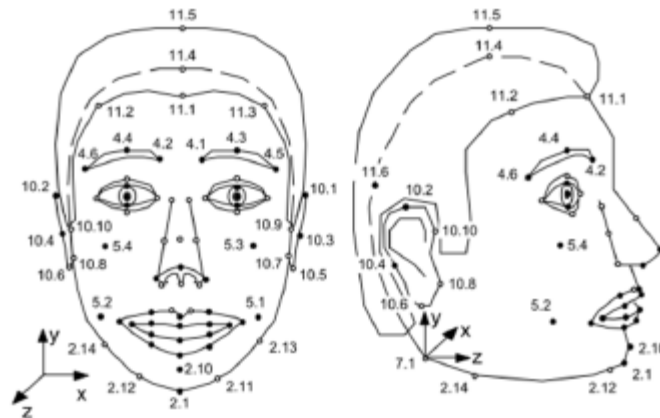
- Componentele unui sistem de recunoaștere a emoțiilor faciale
 - Achiziția imaginilor
 - Detectția feței în imagini
 - Pre-procesarea
 - Determinarea punctelor de interes (*landmark*)
 - Alinierea (*Registration*)
 - Reprezentarea
 - Atribute
 - Reducerea dimensiunii
 - Recunoașterea

Detectarea emoțiilor în expresia feței

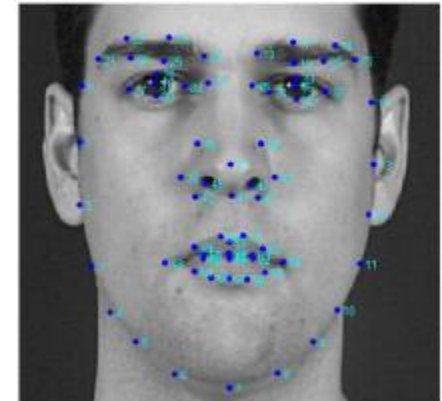
□ Sistem – Pre-procesarea

■ Determinarea punctelor de interes (*landmark*)

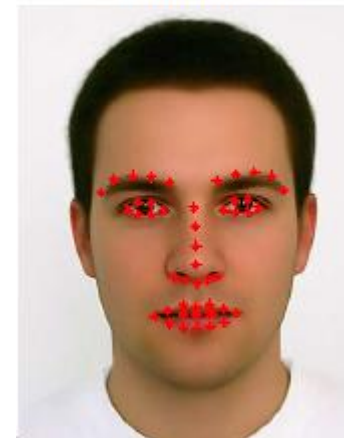
□ Standarde internaționale MPEG-4



68 points model



49 points model



Detectarea emoțiilor în expresia feței

□ Sistem – Pre-procesarea

■ Determinarea punctelor de interes (*landmark*)

□ Standarde internaționale MPEG-4

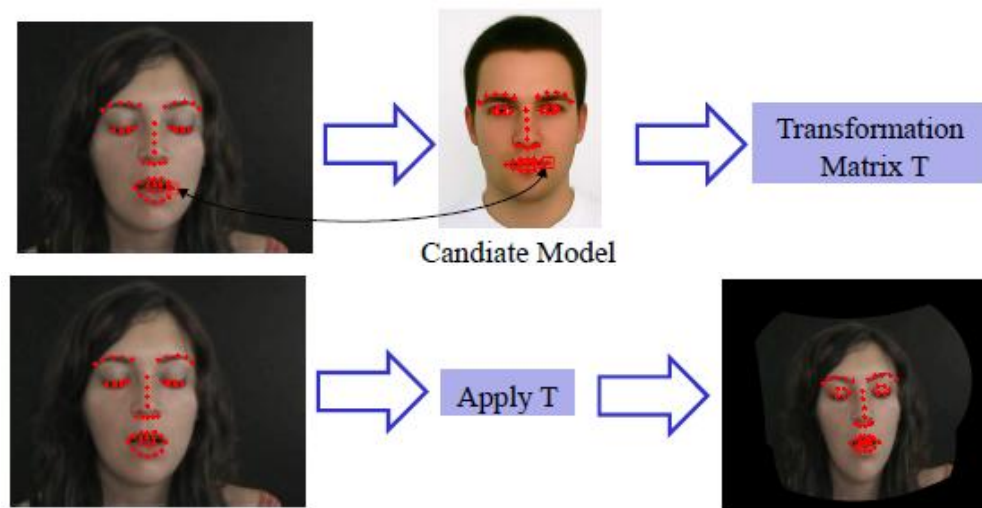
□ Algoritmi

- *Active Shape Model (ASM)*
- *Active Appearance Model (AAM)*
- *Constrained Local Model (CLM)*
- *Incremental Parallel Cascade of Linear Regression Method (Chehra)*
- *Discriminative Response Map Fitting (DRMF)*

Detectarea emoțiilor în expresia feței

□ Sistem – Pre-procesarea

- Determinarea punctelor de interes (*landmark*)
- Alinierea (*registration*)
 - găsirea transformării sau deformarea care minimizează discrepanțele între două sau mai multe obiecte
 - Ideea de bază
 - Calcularea transformării
 - Realizarea transformării afine (scalare, rotație, translație, scalare, reflecție)



Detectarea emoțiilor în expresia feței

- Sistem – Pre-procesarea – Alinierea feței
 - La nivelul întregii fețe
 - Tehnici rigide de aliniere – aliniere globală , la nivelul întregii fețe
 - Detectarea punctelor de interes de pe față
 - Ex. Cei doi ochi
 - Ex. Nas și gură
 - și folosirea lor pentru a calcula o transformare globală: față – prototip
 - Aliniere generică
 - *Active Appearance Model (AAM)*
 - *Robust Fast Fourier Transformation (FFT)*
 - Modelul Lucas-Kanade
 - Tehnici ne-rigide
 - Aliniere locală
 - Tehnica AAM
 - Tehnica SIFT-flow
 - Tehnica *Avatare image registration*
 - La nivelul unei părți din față
 - Numărul, dimensiunea și locația părților poate varia
 - Tehnica AAM
 - Părțile sunt localizate ca regiuni fixe în jurul punctelor de interes
 - La nivelul unui punct de pe față

Detectarea emoțiilor în expresia feței

- Sistem – Reprezentarea feței prin attribute
 - În funcție de frame-urile considerate
 - Attribute spațiale
 - Codifică secvențe de imagini frame-by-frame
 - Attribute spațio-temporale
 - Codifică vecinătăți de frame-uri
 - Tipul de informații codificate
 - Attribute bazate pe aspect
 - Folosesc informații despre textură (intensitatea pixelilor)
 - Attribute bazate pe formă
 - Descriu formele și ignoră textura

Detectarea emoțiilor în expresia feței

□ Sistem – Extragerea atributelor

■ Atribute spațiale

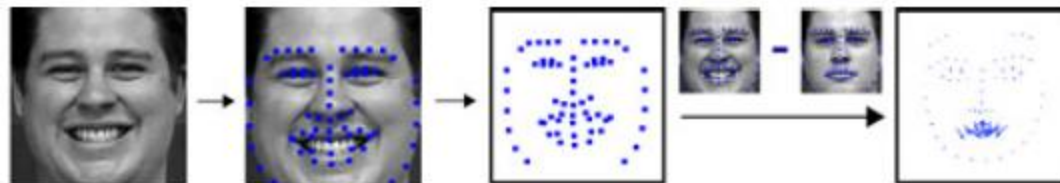
- Reprezentări ale formei
- Reprezentări ale aspectului
- Reprezentări ale părților feței

■ Atribute spațial-temporale

- Reprezentări ale formei
- Reprezentări ale aspectului

Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor – Atribute spațiale – Reprezentări ale formei
 - Puncte de interes
 - Coordonate
 - Distanțe între puncte
 - Distanțe și unghuri între puncte
 - Ochi închiși sau deschiși
 - Starea obrazilor
 - Se folosesc în completarea informațiilor legate de aspect



Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor – Atribute spațiale – Reprezentări ale aspectului
 - De nivel primar
 - Histograme
 - Ideea de bază
 - Se extrag atribute locale și se înglobează într-o imagine transformată
 - Se grupează atributele locale în regiuni uniforme
 - Se colectează atributele fiecărei regiuni în histograme locale
 - Se concatenează histogramele locale
 - Proprietăți
 - Robuste la variații de iluminare
 - Invariante la variații de iluminare globale
 - Se pot normaliza – robuste
 - Robuste la erorile de aliniere
 - Simple de calculat
 - Neajunsuri
 - Afectate negativ de eroarea de identitate (identity bias)
 - Exemple
 - *Local binary pattern* (LBP)
 - *Local phase quantisation* (LPQ)
 - HOG
 - *Quantised local Zernike moments* (QLZM)

Detectarea emoțiilor în expresia feței

- ❑ Sistem – Extragerea atributelor – Attribute spațiale – Reprezentări ale aspectului – De nivel primar – Histograme
 - Local binary pattern (LBP)
 - ❑ Descrie, local, variația texturii într-o regiune circulară
 - ❑ Asociază acestei variații un întreg

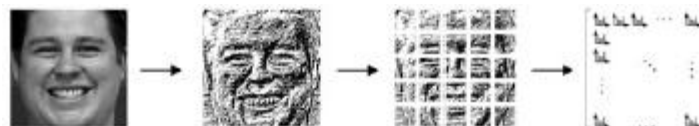
example	thresholded	weights																											
<table><tr><td>6</td><td>5</td><td>2</td></tr><tr><td>7</td><td>6</td><td>1</td></tr><tr><td>9</td><td>8</td><td>7</td></tr></table>	6	5	2	7	6	1	9	8	7	<table><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td></td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	1	0	0	1		0	1	1	1	<table><tr><td>1</td><td>2</td><td>4</td></tr><tr><td>128</td><td></td><td>8</td></tr><tr><td>64</td><td>32</td><td>16</td></tr></table>	1	2	4	128		8	64	32	16
6	5	2																											
7	6	1																											
9	8	7																											
1	0	0																											
1		0																											
1	1	1																											
1	2	4																											
128		8																											
64	32	16																											

Pattern = **11110001**

LBP = $1 + 16 + 32 + 64 + 128 =$ **241**

C = $(6+7+8+9+7)/5 - (5+2+1)/3 =$ **4.7**

- ❑ Histogramele se determină pe baza acestor întregi
 - Dimensiunea reprezentării depinde de câți întregi se folosesc



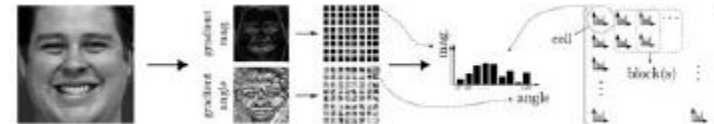
Detectarea emoțiilor în expresia feței

- ❑ Sistem – Extragerea atributelor – Atribute spațiale – Reprezentări ale aspectului – De nivel primar – Histograme

- *Local binary pattern* (LBP)
- *Local phase quantisation* (LPQ)
 - ❑ Similar cu LBP



- HOG



- *Quantised local Zernike moments* (QLZM)
 - ❑ o vecinătate este descrisă prin momente locale Zernike
 - ❑ Fiecare coeficient al momentului descrie variația la o scară și orientare unică – informațiile aduse de diferiți coeficienți ai momentului nu se suprapun
 - ❑ Descriptorul QLZM se obține prin cuantificarea tuturor coeficienților într-un întreg
 - ❑ Histogramamele locale iau în calcul acești întregi

Detectarea emoțiilor în expresia feței

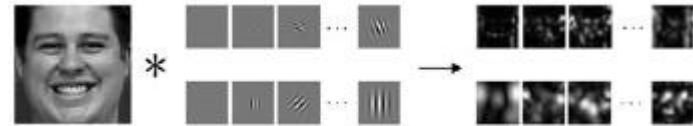
❑ Sistem – Extragerea atributelor – Atribute spațiale – Reprezentări ale aspectului

■ De nivel primar

❑ Histograme

❑ Reprezentări Gabor

- Imaginea este supusă unei convoluții cu filtre Gabor (cu diferite scări și orientări)



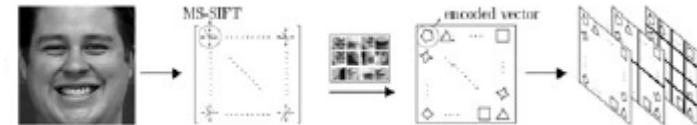
❑ Reprezentări bazate pe date

■ Bag of Word

■ Vecinătăți locale

- Se extrag atribute locale (ex. SIFT) dense din locații fixe

- Se măsoară similaritatea acestor atribute cu o mulțime de atribute cunoscute (visual words) la nivelul unui set de date (visual vocabulary) prin-un algoritm *locality constrained linear coding*



■ De nivel avansat

❑ Factorizări de matrici non-negative

❑ Sparse coding

■ De tip cascadă

❑ Straturi cu informații brute

❑ Straturi cu informații de nivel înalt

Detectarea emoțiilor în expresia feței

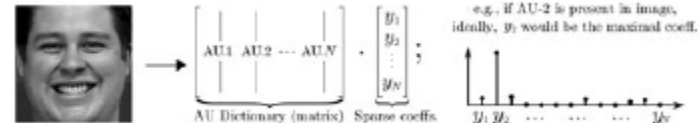
□ Sistem – Extragerea atributelor – Attribute spațiale – Reprezentări ale aspectului

- De nivel primar
- De nivel avansat



□ Factorizări de matrici non-negative

- O matrice (image) este descompusă în mai multe matrici non-negative (imagini de bază)
- Atributele = coeficienții imaginilor de bază
 - Se calculează prin minimizare
- Descompunerea
 - nu e unică
 - Poate avea diferite semnificații
- Ex. Graph-preserving NMF
 - Fața este descompusă în componente spațiale independente
- Ex. Subclass discriminant NMF



□ Sparse coding

- Orice imagine poate fi rară într-un anumit domeniu
- Se poate găsi o transformare a imaginii în care majoritatea coeficienților să fie 0
- Transformarea poate fi
 - Adaptată la date
 - Ne-adaptată (transformată Fourier)

□ Feature learning

Detectarea emoțiilor în expresia feței

- ❑ Sistem – Extragerea atributelor – Atribute spațiale – Reprezentări ale aspectului
 - De nivel primar
 - ❑ Robuste la variațiile de iluminare și erorile de aliniere
 - De nivel avansat
 - ❑ Robuste la erori de identificare
 - ❑ Pot genera atribute interpretabile semantic
 - De tip cascadă (reprezentări ierarhice)
 - ❑ Straturi cu informații brute
 - ❑ Straturi cu informații de nivel înalt
 - ❑ Deep learning
 - Se învață reprezentări ierarhice pe mai multe nivele
 - 2 nivele cu informații brute
 - Imaginile suportă o convoluție cu filtre locale (smooth – diferențe locale) învățate din date
 - Rezultatele convoluției sunt agregate prin diferite operații (ex. pooling)
 - Alte nivele cu informații de nivel avansat pentru diferite scopuri
 - Identificarea elementelor ocluzionate

Detectarea emoțiilor în expresia feței

□ Sistem – Extragerea atributelor – Attribute spațiale – Reprezentări ale părților feței

- Calculate la nivelul unei părți
- Se ignoră relațiile spațiale între părți
- Folosite în recunoașterea emoțiilor spontane

■ Ex.

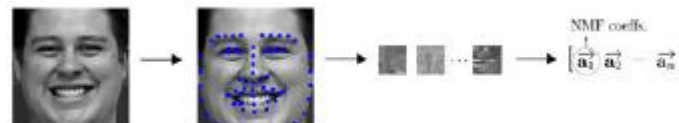
□ Part-based SIFT

- Se calculează descriptori SIFT (cu scară și orientare fixă) pentru fiecare parte a feței



□ Part-based NMF

- Se elimină detaliile de textură personale
- Se descompune imaginea părții în NM

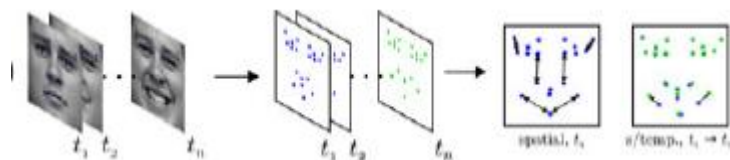


Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor
 - Atribute spațiale
 - Atribute spațio-temporale
 - Reprezentări ale formei
 - Geometric Features from Tracked Facial Points
 - Reprezentări ale aspectului
 - Atribute de nivel primar din planurile ortogonale
 - Convoluții cu filtre netede (smooth filter)
 - Reprezentări Haar spațio-temporale
 - Reprezentări ale deformărilor

Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor – Atribute spațio-temporale – Reprezentări ale formei
 - Geometric Features from Tracked Facial Points
 - Analiza temporală a variației activității musculare
 - Ideea:
 - Se folosesc
 - coordonatele fiecărui punct de interes
 - Lungimea și unghiul liniile care conectează toate aceste puncte (luate în perechi)
 - Se stabilesc diferențele față de o față neutră



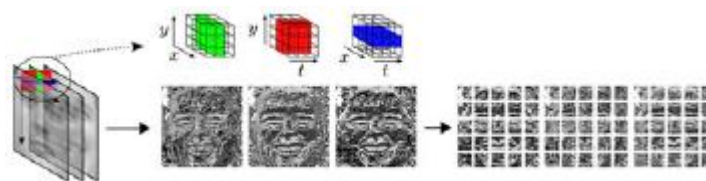
Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor – Attribute spațio-temporale – Reprezentări ale aspectului
 - Attribute de nivel primar din planurile ortogonale
 - Convoluții cu filtre netede (smooth filter)
 - Reprezentări Haar spațio-temporale
 - Reprezentări ale deformărilor
 - Bag of Words temporal

Detectarea emoțiilor în expresia feței

- Sistem – Extragerea atributelor – Atribute spațio-temporale – Reprezentări ale aspectului
 - Atribute de nivel primar din planurile ortogonale
 - Three orthogonal planes (TOP)
 - Se extrag atribute din vecinătăți spațio-temporale în 3 planuri:
 - Planul spațial (x-y – LBP)
 - Planul vertical spațio-temporal (y-t)
 - Planul orizontal spațio-temporal (x-t)
 - Se determină histograme locale ale regiunilor spațio-temporale

□ LBP-TOP



□ LPQ-TOP



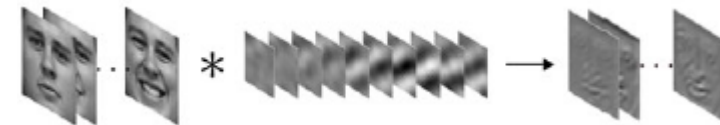
Detectarea emoțiilor în expresia feței

□ Sistem – Extragerea atributelor – Atribute spațio-temporale – Reprezentări ale aspectului

■ Convoluții cu filtre netede (smooth filter)

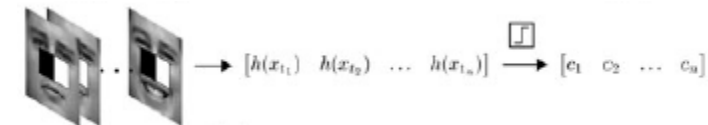
□ Filtre spațio-temporale Gabor

□ Filtrare cu componente spațio-temporale independente (IC)



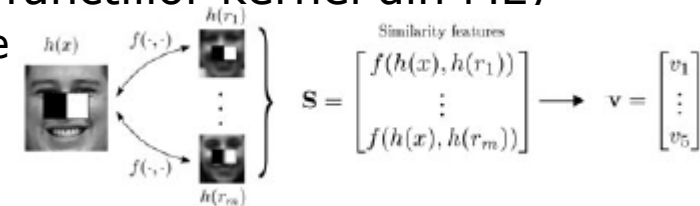
■ Reprezentări Haar spațio-temporale

□ Atribute Haar dinamice



□ Atribute de similaritate (asemănător funcțiilor kernel din ML)

■ Se aplică un filtru Haar fiecărei frame

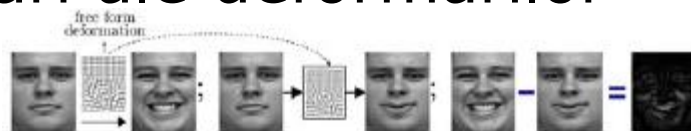


- se reține într-un vector valoarea unei funcții de similaritate între rezultatul filtrării și outputul corespunzător unei imagini referință
- Se calculează histograma vectorilor aferenți întregii secvențe de frame-uri

Detectarea emoțiilor în expresia feței

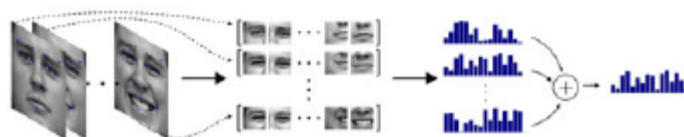
- ❑ Sistem – Extragerea atributelor – Atribute spațio-temporale – Reprezentări ale aspectului

- Reprezentări ale deformărilor



- Bag of Words temporal

- ❑ Fiecare frame e reprezentat prin part-based SIFT și comprimat prin PCA, rezultând un vector
- ❑ Vectorul se codează în stil BoW
- ❑ Toți vectorii se "reunesc" într-o histogramă



Detectarea emoțiilor în expresia feței

- ❑ Sistem – Reducerea dimensiunii de reprezentare
 - Pooling
 - ❑ Scop
 - Agregarea atributelor locale
 - ❑ Tehnici
 - Sumă
 - Histogramă
 - Maxim
 - Norma L_2
 - Stochastic
 - Maxout
 - Selectarea atributelor (*Feature selection*)
 - ❑ Scop
 - Selectare & Ponderare
 - ❑ Tehnici
 - Supervizate - Boosting (AdaBoost, GentleBoost)
 - Extragerea atributelor (*Feature extraction*)
 - ❑ Se calculează noi attribute
 - ❑ Transformare
 - Ne-adaptată
 - Transformarea discretă cosine (DCT)
 - LDA
 - Adaptată
 - PCA

Detectarea emoțiilor în expresia feței

□ Sistem – recunoașterea

■ Scop

□ Recunoașterea etichetei unei

- Acțiuni
- Emoții

□ Recunoașterea intensității unei

- Acțiuni
- Emoții

□ Determinarea etapei producerii

- Unei AU
- Unei combinații de AU

■ Algoritmi

□ Statistici

- Doar date spațiale
 - SVM
- Date temporale
 - HMM&SVM/Boosting
 - Dynamic bayesian network
 - CRF

Detectarea emoțiilor în expresia feței

□ Dezvoltări

- Noi reprezentări bazate pe forme
 - Viziunea umană – percepția emoțiilor de pe față
 - Forme continue, nu doar discrete
- Noi reprezentări bazate pe părți ale feței
- Noi reprezentări ierarhice
- Noi tehnici de aliniere (spațio-temporală)
- Date 3D

Detectarea emoțiilor în gesturi

□ Gesturi

- Mimica feței
- Gesturi ale corpului

□ Emoțiile faciale

- Științele cognitive
 - Ce informație trebuie reprezentată?
- Științele computaționale
 - Cum trebuie reprezentată (codată) informația?

□ Percepția feței

- Abordarea holistică
 - Fața este modelată ca un întreg, fără părți (componente) care ar putea fi izolate
- Abordarea bazată pe componente
 - Anumite attribute ale feței pot fi procesate individual
- Abordarea bazată pe configurație
 - Se modelează relațiile spațiale dintre componentele feței (ex. ochiul stâng-ochiul drept, nas-gură)

Recunoașterea gesturilor

- Gesturi – concept & clasificare
- Sisteme inteligente

Recunoașterea gesturilor

□ Gesturi

■ Postură (poziție)

- Configurația degetelor (fără mișcarea mâinii)

■ Gest

- Mișcare dinamică a mâinii, cu sau fără mișcarea degetelor

■ Gesticulare

- Mișcare spontană a mâinii/brațelor care însoțește vorbirea
- 90% din gesturile umane

■ Pantomimă

- Mișcări ale mâinii/brațelor/corpului fără vorbire

Recunoașterea gesturilor

□ Gesturi

■ În funcție de suport

- Bazate pe caracteristici observabile (statice sau dinamice)
 - Statice (independente de timp) – poziția mâinii e constantă de-a lungul gestului
 - Dinamice (dependente de timp)
- Bazate pe interpretare (ilustrative)

■ În funcție de modul de implicare al executantului

□ Conștiente

- Regulare - pentru controlarea interacțiunii
 - Ridicarea mâinii pentru a solicita cuvântul
- Afective
 - Retragera din fața unui obiect poate indica frica
- **Ilustrative**
 - Coezive (unitare)
 - Metaforice - abstractizări
 - Contextuale – identificarea unei persoane sau a unei perioade de timp sau a unei direcții
 - Reprezentative – cățărul în copaci
 - Bătăi – scurte și rapide
- **Emblematic** (autonome) -- cultural specificeș pot înlocui cuvinte/expresii
 - Ex. Pa! Pa!, aprobarea din cap, thumbs-up!

□ Inconștiente

- Adaptive –
 - Ex. Zgâlțâirea capului, bâțâitul din picior, aranjarea ochelarilor

■ In funcție de localizarea spațio-temporală

- Gesturi temporale
- Gesturi spațiale
- Gesturi spațio-temporale

Recunoașterea gesturilor

□ Sistem – pași

- Achiziția datelor
- Modelarea gesturilor
 - Localizarea (segmentarea) mâinii
 - Filtrări și eliminarea zgomotului
 - Detecția conturului
 - Normalizarea
- Extragerea atributelor
- Recunoașterea gesturilor

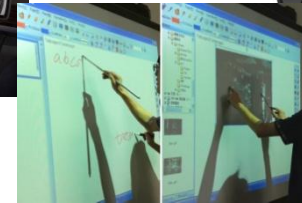
Recunoașterea gesturilor

□ Sistem – Achiziția datelor

■ Dispozitive

□ Bazate pe contact

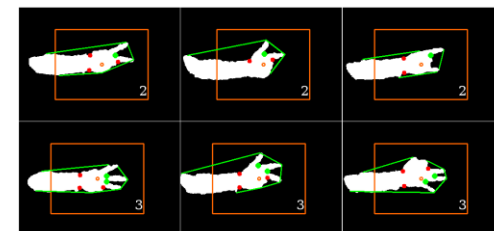
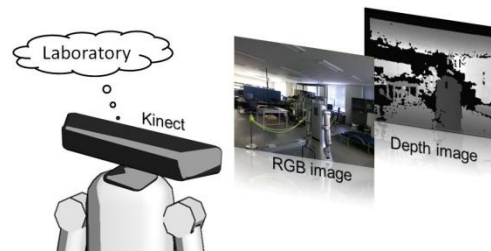
- Mănuși,
- Accelerometre,
- Ecrane multi-touch



<https://www.youtube.com/watch?v=kXuxK6IeQfo>

□ Bazate pe dispozitive de vizualizare

- Camere
 - RGB
 - RGB-D
- Markere



Recunoașterea gesturilor

■ Sistem – Achiziția datelor

- Baze de date – Pramod Kumar Pisharady and Martin Saerbeck , Recent methods and databases in vision-based hand gesture recognition: A review, Computer Vision and Image Understanding, Vol. 141, 2015:152-165

No.	Name, Year	Source
1	ChalLearn gesture* data**, 2011	http://gesture.challearn.org/data
2	MSRC-12 Kinect gesture* dataset**, 2012	http://research.microsoft.com/en-us/um/cambridge/projects/msrc12/
3	ChalLearn multi-modal gesture data**, 2013	http://sunat.uoc.edu/challearn/
4	NUS hand posture dataset-II, 2012	http://www.cse.nus.edu.sg/sfpag/ekpv/NUS-HandSet/
5	6D motion gesture database*, 2011	http://www.cse.gatech.edu/6DMC/6DMC.html
6	Sebenien Marcel Interact play database, 2004	http://www.idiap.ch/resource/interactplay/
7	NATOPS alerting handling signals database*, 2011	http://groups.csail.mit.edu/mug/natops/
8	Sebenien Marcel hand posture and gesture dataset, 2001	http://www.idiap.ch/resource/gestures/
9	Gesture dataset by Shen et al., 2012	http://users.oecs.northwestern.edu/~xsh835/GestureDataset.zip
10	Gesture dataset by Yoon et al., 2001	available on e-mail request to yoonh@setfire.kr
11	ChAIR-Gest multi-modal dataset**, 2013	https://project.ula.fr/chairgest/Pages/Download.aspx
12	Sheffield Kinect Gesture Dataset**, 2013	http://shah.maff.shef.ac.uk/data/SheffieldKinectGesture.htm
13	Kock gesture dataset, 2009	http://www.uniacs.usd.edu/~zhuolin/Kockgesturedataset.html
14	NUS hand posture dataset-I, 2010	http://www.cse.nus.edu.sg/sfpag/ekpv/NUS-HandSet/
15	Cambridge hand gesture dataset, 2007	http://www.iis.oe.ic.ac.uk/~tkk/gas_db.htm
16	Posture dataset by Ren et al**, 2011	http://eeewebe.ntu.edu.sg/computervision/people/hong/renzhou/HandGesture.htm
17	ColorTip dataset**, 2013	https://isatga.upc.edu/web/res/colortip
18	NYU Hand Pose Dataset**, 2014	http://cims.nyu.edu/~tompson/NYU_Hand_Pose_Dataset.htm#overview
19	General-HA NDS dataset**, 2014	http://wildhog.ics.utd.edu/9090
20	VPU Hand Gesture dataset (HGDs), 2008	http://www.vpu.epa.uam.es/DS/HGDs/
21	Dataset by Kawulok et al., 2014	http://sun.aol.polis.pl/~kawulok/gestures/
22	ASL Finger Spelling Dataset**, 2011	http://personal.ee.surrey.ac.uk/Personal/W.Pugsault/index.php?section=FingerSpellingDataset

*All the gestures in this dataset are not hand gestures. Some are body gestures.

**These are RGB-D sensor based datasets, containing depth/ skeletal information.

No.	Description	S/D	Works
1	ChalLearn Gesture Challenge, 62,000 samples	D	[82, 90-92, 100, 127]
2	12 classes, 30 subjects, 6,244 samples	D	[126]
3	20 classes, 27 subjects, 13,858 samples	D	[128]
4	10 classes, 40 subjects, 2,750 samples, complex background	S	[44, 129]
5	20 classes, 28 subjects, 5,600 samples	D	[130]
6	16 classes, 22 subjects, 50 samples/ subject	D	[13, 131]
7	24 classes, 20 subjects, 9,600 samples	S & D	[132]
8	Three hand posture datasets, with 10 (gray scale), 12 (color), and 6 (gray scale) classes. One hand gesture dataset with 4 classes	S & D	[63-65, 133]
9	10 classes, 15 subjects, 1,050 samples	S & D	[35]
10	48 classes, 20 subjects, 9,600 samples	D	[11]
11	10 classes, 10 subjects, 1,200 samples recorded with Kinect and inertial motion units	D	[134]
12	10 classes, 6 subjects, 2,160 samples recorded with Kinect and RGB cameras	D	[135]
13	14 classes, 3 subjects, 126 training and 168 testing samples	D	[136]
14	10 classes, 1 subject, 240 samples, color as well as grey scale	S	[45]
15	9 classes, 2 subjects, 900 image sequences, with different illumination conditions	S & D	[137]
16	10 classes, 10 subjects, 1000 samples, color as well as depth maps, cluttered background	S	[88]
17	7 subjects, 9 classes, 7 training sequences of between 600-2000 depth frames	S	[118]
18	2 users, data from 3 Kinects (frontal and 2 sides), 72757 and 8252 frames in training and test sets	S	[138]
19	22 sequences, different view-points, scales, poses, and occlusions	S	-
20	12 classes, 11 subjects, 1 video per gesture (252 frames)	S	[139]
21	32 classes, 18 subjects, gestures from Polish Sign Language and American Sign Language (ASL)	S	[140]
22	24 classes, 9 subjects, 65,000 samples	S	[141]

S-static, D-dynamic

Recunoașterea gesturilor

□ Sistem – Modelarea gesturilor

■ Localizarea (segmentarea) mâinii

- Metoda pragului (thresholding)
 - Range/depth thresholding
 - Colour thresholding
 - Speed thresholding
 - Otsu thresholding (bazată pe histogramă)
- Metode bazate pe culoarea pielii (skin-based)
- *Substraction*
 - Are nevoie de o imagine referință
- Modele statistice
 - *Bayesian rule based*
 - *Gaussian mixture model*
 - *Expectation Maximisation*
- Normalizarea culorii

■ Filtrări și eliminarea zgomotului

■ Detecția conturului

■ Normalizarea

Recunoașterea gesturilor

□ Sistem –Modelarea gesturilor

■ Localizarea (segmentarea) mâinii

■ Filtrări și eliminarea zgomotului

□ Zgomotul Sare și piper – prin filtre

- median
- morfologice
- contra armonice medii

□ Operații morfologice

- Eroziunea – tinde să reducă dimensiunile caracteristicilor luminoase ale imaginii prin corelare cu zonele întunecate adiacente
- Diluarea – inversul eroziunii

□ Zgomotul muchiilor (contururilor)

- Filtru mediu multi-dimensional

■ Detecția conturului

■ Normalizarea

Recunoașterea gesturilor

□ Sistem – Modelarea gesturilor

- Localizarea (segmentarea) mâinii
- Filtrări și eliminarea zgomotului
- Detecția conturului

□ In functie de derivări

- Metode bazate pe gradient – max&min a derivatei funcției de intensitate de ordin I
- Metode bazate pe Laplacian – trecerile prin 0 a derivatei funcției de intensitate de ordin II

□ Tehnici

- Sobel
- Robert
- Prewitt
- Laplacian of Gaussian
- Canny

■ Normalizarea

Recunoașterea gesturilor

- Sistem – Modelarea gesturilor
 - Localizarea (segmentarea) mâinii
 - Filtrări și eliminarea zgomotului
 - Detecția conturului
 - Normalizarea (reducerea spațiului atributelor)
 - Decuparea (cropping)
 - Redimensionarea
 - Filtre Gabor
 - Modificări de muchii și de textură
 - Tehnici de ML

Recunoașterea gesturilor

- Sistem – pași
 - Achiziția datelor
 - Modelarea gesturilor
 - Localizarea (segmentarea) mâinii
 - Filtrări și eliminarea zgomotului
 - Detecția conturului
 - Normalizarea
 - **Extragerea atributelor**
 - **Recunoașterea gesturilor**

Recunoașterea gesturilor

□ Sistem – Extragerea atributelor

■ 2D

□ Șabloane deformabile



□ Atribute de formă

■ Atribute geometrice

- locația și poziția degetelor și a palmei

■ Atribute non-geometrice (blind features)

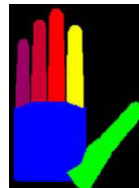
- Culoarea, profilul&textura, conturul, momentele imaginii, vectorii caracteristici (Eigen) (14,16,17)



□ Mișcare



□ Markere de culoare



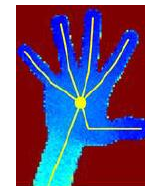
■ 3D

□ Volumetric

□ Profil – structura mâinii

□ Geometrice – animație

- Polygon meshes and cardboard models



Recunoașterea gesturilor

□ Sistemele de recunoaștere a gesturilor

■ Detecție

- Culoarea pielii
- Formă
- Textură
- Modele 3D ale mâinii
- Mișcarea mâinii

■ Tracking

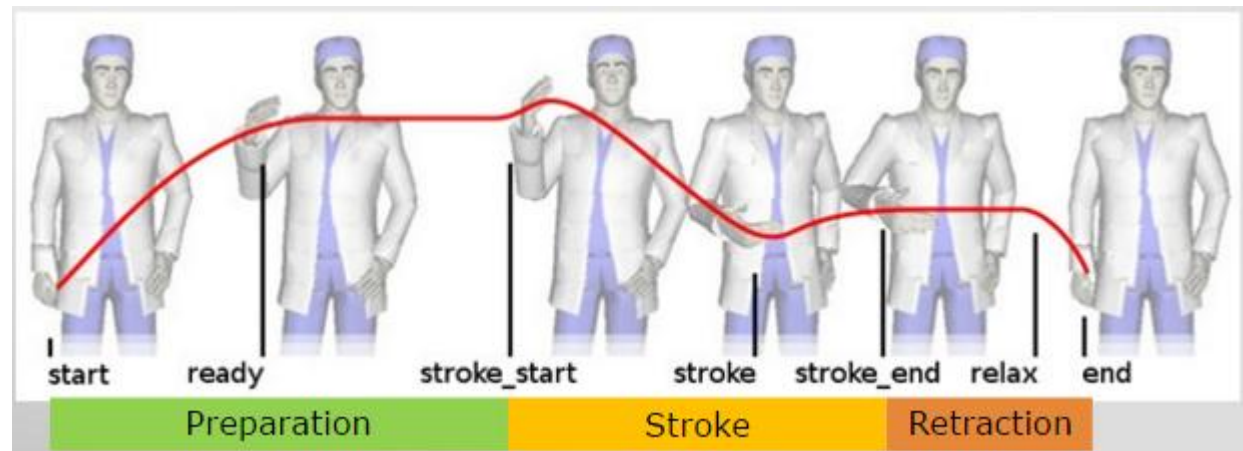
- Metode bazate pe șabloane
 - *Correlation-based feature tracking*
 - *Contour based tracking*
- Metode de estimare
 - Filtre Kalman
 - Filtrări de particule
- CamShift

■ Recunoaștere

- Metode nesupervizate
 - K-Means
 - K-nearest neighbour
 - Mean shift
- Metode supervizate
 - SVM
 - HMM
 - *Dynamic time warping* (alinierea semnalelor)
 - *Time delay neural networks*
 - *Finite state machine*

Recunoașterea gesturilor

- Sistem – recunoașterea
 - Metode statice
 - Metode bazate pe mișcare
 - Etape ale prelucrării temporale
 - Pregătirea
 - Nucleul
 - Retragera



Recunoașterea gesturilor

□ Sistem – recunoașterea

■ Metode bazate pe aspect (*view*)

- Folosesc attribute extrase din imagini de antrenament pentru a modela aspectul
- Compară acești parametrii pentru o imagine de test

□ Avantaje

- Performanțe (temporale) ridicate (real-time)

□ Dezavantaje

- Dependente de view (unghiul de vizualizare)

■ Metode bazate pe modele 3D ale mâinii

- Modele cinetice 3D care estimează parametrii unghiulari și liniari ai modelului mâinii
- Fără antrenare!!!!

□ Avantaje

- Acoperă toate perspectivele

□ Dezavantaje

- Necesită multe imagini pentru a acoperi toate caracteristicile de formă și perspectivele unui obiect

Recunoașterea gesturilor

□ Sistem – recunoașterea

■ Metode bazate pe aspect (view)

□ Cele mai cunoscute modele

■ Model bazat pe culoare

- Atribute de culoare multi-scale
- Modele ierarhice
- Filtrări de particule



■ Model geometric al siluetei

- Perimetrul, convexitatea, suprafața, *bounding box/ellipse, elongation, rectangularity, centroid and orientation*



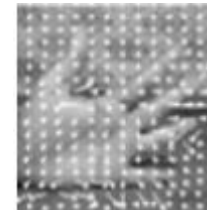
■ Model Gabarit deformabil

- Contururi active deformabile



■ Model bazat pe mișcare

- Histograme locale de mișcare



Recunoașterea gesturilor

□ Sistem – recunoașterea

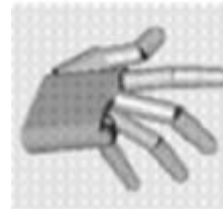
■ Metode bazate pe modele 3D ale mâinii

□ Cele mai cunoscute modele

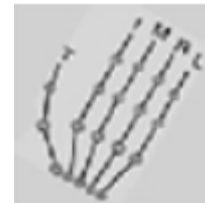
- Model 3D texturat volumetric
 - Profilul mâinii și textura pielii



- Model 3D geometric



- Model 3D de profil



Recunoașterea gesturilor

□ Sistem – recunoașterea

■ Camere RGB

□ Recunoașterea gesturilor dinamice

- HMM și alte modele statistice
- RNA și alte metode de ML
- Metode bazate pe vectori proprii (Eigen)
- Curve fitting
- Programare dinamică / Dynamic time wrapping

□ Recunoașterea poziției mâinii

- Metode de învățare supervizate
- Metode de învățare ne-supervizate
- Potrivire de grafe
- Metode bazate pe modele 3D

Recunoașterea gesturilor

□ Recunoașterea gesturilor dinamice

- Pramod Kumar Pisharady and Martin Saerbeck , Recent methods and databases in vision-based hand gesture recognition: A review, Computer Vision and Image Understanding, Vol. 141, 2015:152-165

Work	Features used	Classification method	Application
[10]	direction of hand movement	HMM	browsing commands in <i>PowerPoint</i> ^(R) presentation
[11]	hand location, angle & velocity	HMM	HCI- recognizing alphanumeric characters & graphic elements
[14]	Fourier descriptors/ optical flow	HMM	Taiwanese sign language
[12]	hand shape & hand motion	HMM	remote robot control
[17]	3D articulation data	accumulative HMM	controlling lights and curtains in smart home
[13]	3D trajectory, hand displacement, color & shape of hand blob	HMM & IOHMM	interact-play, manipulation
[25]	haar-like features	statistical/ syntactic anal.	not specified
[24]	directional features	DBN	controlling media player
[39]	3D motion trajectory	curve fitting	3D bioinformatics data visualization navigation
[37]	hand shape / trajectory	predictive eigen tracker	audio player control
[32]	2D motion field / trajectory	NN	American sign language
[34]	Fourier descriptors (shape of hand blob)	RBF, HMM & RNN	manipulation of objects in windows user interface
[29]	hand motion (motion energy)	FSM	HRI
[42]	3D hand motion features	CDFD & Q-DFFM	Dutch sign language

Descriptions: HMM-hidden Markov model, IOHMM-input / output hidden markov model, HCI

-human computer interaction, DBN-dynamic Bayesian network, NN-neural network, RBF-radial basis function.

RNN-recurrent neural networks, FSM- finite state machines, HRI-human robot interaction, CDFD-combined

discriminative feature detectors, Q-DFFM-quadratic classification on discriminative features fisher mapping

Recunoașterea gesturilor

Work	Accuracy	Class	Subj.	Samp.	UI.	Spot	BG	Noise	Scale	Light	Exten.	CV	Data
[10]	93.14	10	8	6.2	0	1	0	0	0	0	0	0	0
[11]	93.25	48	20	5	0	1	0	0	1	1	0	0	0
[14]	93.6	20	20	3	0	0	1	0	1	0	0	0	0
[12]	81.71	5	5	14	0	1	1	1	0	1	0	0	0
[17]	95.42	8	1	60	0	1	0	0	0	0	0	0	0
[13]	75 & 98	16 & 7	20 & 7	50 & 10	0	0	0	0	0	0	0	0	1
[25]	87.21	4	1	25	0	1	0	0	1	1	0	0	0
[24]	99.59	10	7	1	0	1	0	0	0	0	0	1	0
[39]	97.9	10	4	2.38	0	0	1	0	0	1	0	0	0
[37]	100	8	1	2	0	0	0	0	1	0	0	0	0
[32]	96.21	40	1	7.6	0	1	1		0	1	0	1	0
[34]	91.9	14	1	21.07	0	1	0	0	1	0	0	0	0
[29]	not reported	5	1	1	0	0	0	0	0	0	0	0	0
[42]	92.3	120	75	15	1	1	0	1	0	0	0	1	0

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Scale**-Variation in scale/ size considered or not, **Light**-Variation in lighting considered or not, **Exten.**-Online or offline learning, 1 for online, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public

Recunoașterea gesturilor

□ Recunoașterea poziției mâinii

Work	Features	Classification method	Application
[61]	geometric distance	DLLE / PNN	manipulation of objects in windows user interface
[52]	intrinsic geometry of hand	SLLE	Chinese sign language
[51]	multivalued features (centroid, compactness, area of hand)	RIL	gesture commands
[50]	discrete Fourier transform based distance metric	nearest neighbour / maximum likelihood	gesture commands
[44]	shape, texture and color features	SVM	recognition against complex backgrounds
[63]	Gabor jets	EGM	HRI
[69]	joint angles	3D model fitting	not specified
[45]	shape and texture features	Fuzzy-Rough classifier	HRI
[62]	shape features	EGM	not specified
[58]	Gabor features	SVM	recognition under varying illumination
[67]	Histogram of Oriented Gradient	EGM	not specified

Descriptions: PNN-probabilistic neural network, DLLE-distributed locally linear embedding, SLLE-supervised locally linear embedding, RIL-recursive induction learning, SVM-support vector machines, EGM-elastic graph matching. HRI-human robot interaction

Recunoașterea gesturilor

Work	Accuracy	Class	Subj.	Samp.	UI	Spot	BG	Noise	Scale	Light	Exten.	CV	Data
[61]	93.2	14	1	20	0	1	0	0	1	1	0	0	0
[52]	90.6	30	1	55	0	1	0	0	0	1	0	0	0
[51]	94.4	20	1	45.4	0	1	0	0	0	0	0	0	0
[50]	98.5	9	4	44.44	1	0	1	0	0	0	1	0	0
[44]	94.36	10	40	5	1	0	1	1	1	0	0	1	0
[63]	85.8	12	19	1.48	1	0	1	0	0	0	0	1	1
[69]	not reported	4	1	1	0	0	0	0	0	0	0	0	0
[45]	98.75	10	19	18	1	0	0	0	1	1	0	1	1
[62]	96.35	10	19	2.52	1	0	0	0	0	0	0	0	1
[58]	96.1	11	10	6	0	0	1	0	1	1	0	0	0
[67]	99.85	10	24	2.7	1	0	1	0	0	0	0	1	1

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Scale**-Variation in scale/ size considered or not,**Light**-Variation in lighting considered or not, **Exten.**-Online or offline learning, 1 for online, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public

Recunoașterea gesturilor

□ Sistem – recunoașterea

■ Camere RGB-D

- Metode pentru recunoașterea gesturilor dinamice
 - One-shot-learning
 - Extended-Motion-History-Image (Extended-MHI)
 - Coeficient de corelație maxim
 - Cel mai apropiat vecin
 - Dynamic time wrapping
- Metode pentru recunoașterea poziției mâinii
 - Random decision forests
 - Clasificator probabilistic bazat pe distanța de editare

Recunoașterea gesturilor

- Metode pentru recunoașterea gesturilor dinamice și a poziției mâinii

Work	S/D	Sensor	Features	Classification method
[86]	D	CSEM Swiss-ranger SR-2	motion primitives	probabilistic edit distance classifier
[87]	S	ToF and RGB camera	Haarlets	NeN
[88]	S	Kinect	hand/finger shape	template matching using FEMD
[89]	D	Kinect	hand area	classifier based on topology
[90]	D	Kinect	Extended-Motion-History-Image	maximum correlation coefficient
[91]	S	Kinect	depth pixel values	randomized classification forests & voting
[92]	D	Kinect	underlying geometry	least squares fitting
[93]	D	Kinect	Euclidean and log-Euclidean distance	NeN
[94]	D	PrimeSense 3-D camera	probabilistic 2D templates from trajectory	MPLCS classifier
[95]	D	Kinect	spatial and motion features	conditional density propagation
[96]	D	Kinect	position, angle, and direction features	probability; pairwise coupling
[97]	D	Kinect	conditional distance	dynamic time warping

S-static, D-dynamic, ToF-time of flight, NeN-nearest neighbor, FEMD-finger earth mover's distance, MPLCS-most probable longest common subsequence

Recunoașterea gesturilor

Work	Accuracy	Class	Subj.	Samp.	UI	Spot	BG	Noise	Scale	Light	Exten.	CV	Data
[86]	92.9	4	10	3	0	1	0	0	0	0	0	0	0
[87]	99.54	6	1	29.17	0	1	1	1	0	0	0	0	0
[88]	93.9	10	10	1	0	1	1	0	1	1	0	0	0
[89]	not reported	9	1	1	0	1	0	0	0	0	0	0	0
[90]	not reported	8-15	multiple (ChaLearn)	multiple (ChaLearn)	1	1	1	1	1	1	0	0	1
[91]	84.3 & 74.3	24 & 9	4 & 5	100 & 10	1	0	0	0	0	1	0	1	1
[92]	91.7	9	2	80	1	0	0	0	0	1	0	1	1
[93]	99.75	8	20	5	1	0	0	0	1	0	0	1	0
[94]	98.7	10	8	5	1	1	0	0	1	0	0	1	0
[95]	95.9	4	4	1	1	1	1	0	0	0	0	1	0
[96]	97.26	10	6	1	1	1	1	0	1	0	0	1	0
[97]	82	179	18	multiple (ChaLearn)	1	0	1	1	1	1	0	1	1

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Exten.**-Online or offline learning, 1 for online, **Scale**-Variation in scale/ size considered or not, **Light**-Variation in lighting considered or not, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public