

Concepțe și aplicații în Vederea Artificială

Bogdan Alexe

bogdan.alexe@fmi.unibuc.ro

Radu Ionescu

radu.ionescu@fmi.unibuc.ro

Curs optional
semestrul I, 2023-2024

Termen limită

Termene limită

Prima fază - trimiterea codului. Încărcați o arhiva zip cu codul soluției voastre și un fișier pdf ce descrie soluția voastră până marți, 5 decembrie, ora 23:59 la link-ul acesta <https://tinyurl.com/CAVA-2023-TEMA1-SOLUTII>. Includeți în arhiva zip NU-MAI cod (fișiere .py sau .ipynb) sau alte fișiere necesare rulării codului (fișiere cu modelele voastre antrenate, alte imagini auxiliare folosite de voi etc.). Nu includeți în arhiva voastră imaginile inițiale (le avem și noi!!!). **ATENȚIE: nu vom accepta teme trimise după data limită.**

Codul vostru ar trebui să includă un fișier README (vedeți exemplul din materiale) cu următoarele informații: (i) librăriile folosite de voi și necesare pentru rularea soluției voastre; (ii) indicații despre cum ar trebui rulat codul pentru fiecare task. Studenții care nu încarcă un fișier pdf cu descrierea soluției lor vor primi 0 puncte la partea de documentație.

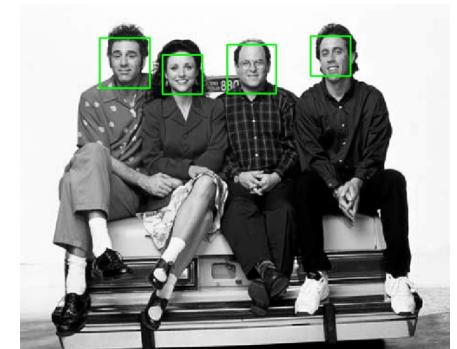
A doua fază - trimiterea rezultatelor. Miercuri, 6 decembrie, vom publica datele de test în directorul *testare* de la adresa <https://tinyurl.com/CAVA-2023-TEMA1>. Veți rula soluția voastră pe imaginile de test și veți încărca rezultatele în aceeași zi ca o arhivă zip folosind următorul link <https://tinyurl.com/CAVA-2023-TEMA1-REZULTATE>.

Cursul trecut

- Localizarea claselor de obiecte
 - metoda ferestrei glisante



- Detectare facială folosind metoda glisării ferestrei și histograme de gradienți orientați



Cursul de azi

- Modelul Bag of Visual Words
 - privire de ansamblu
 - descriptorul SIFT
 - clusterizarea k-means
 - piramida spațială



- Recunoașterea obiectelor pe bază de textură



Modelul bag of visual words



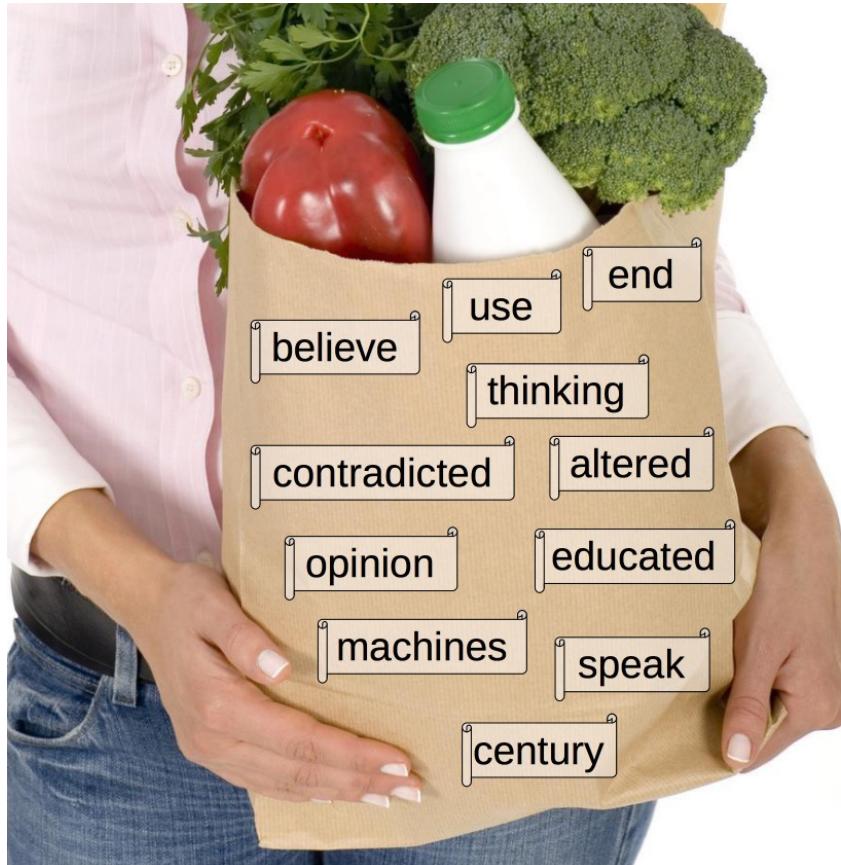
Modelul bag of words utilizat în procesarea limbajului natural

Modelul bag of words (“colecție de cuvinte”) funcționează astfel:

- 1) Fiind dat un set de documente text, se construiește un vocabular cu toate cuvintele care apar în aceste documente
- 2) Reprezentarea bag of words a fiecărui document este obținută prin numărarea cuvintelor din vocabular care apar în document și construcția histogramei (normalizează) de frecvență

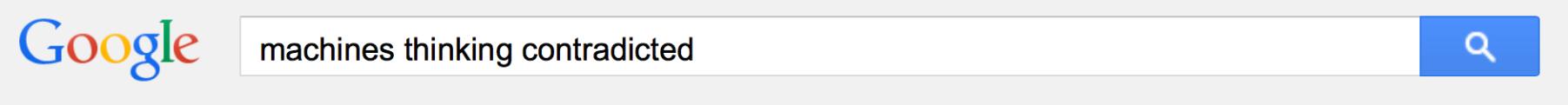
Modelul bag of words utilizat în procesarea limbajului natural

“I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.” – Alan Turing



Modelul bag of words utilizat în procesarea limbajului natural

- Pentru o interogare pe Google:

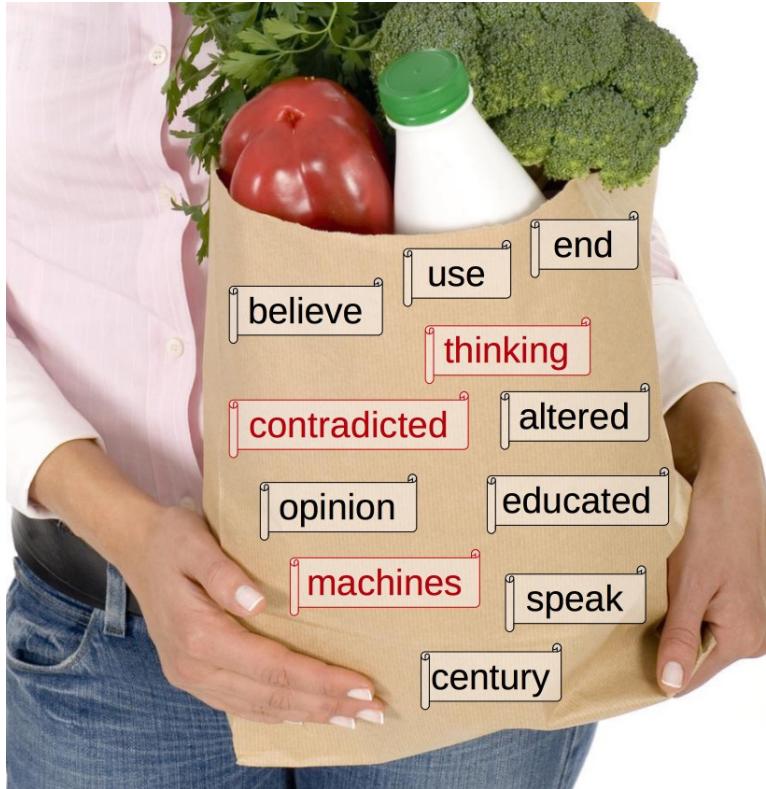


Modelul bag of words utilizat în procesarea limbajului natural

- Pentru o interogare pe Google:

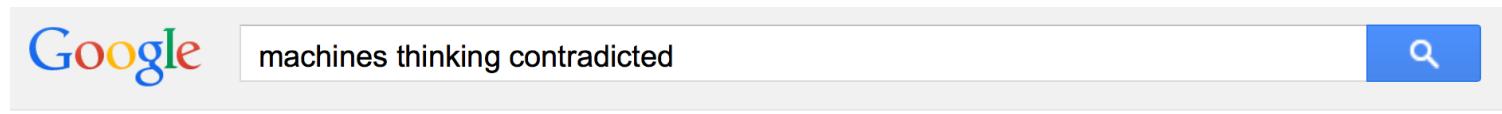


- Modelul bag of words poate găsi imediat documentele care conțin termenii căutați:



Modelul bag of words utilizat în procesarea limbajului natural

- Întorcând astfel documentele relevante:



About 15,100,000 results (0.29 seconds)

Alan Turing - Wikiquote

https://en.wikiquote.org/wiki/Alan_Turing ▾

... much that one will be able to speak of **machines thinking** without expecting to be **contradicted**. ... These questions replace our original, "Can **machines think**?".

Computing Machinery and Intelligence A.M. Turing

www.loebner.net/Prizef/TuringArticle.html ▾

I propose to consider the question, "Can **machines think**? so much that one will be able to speak of **machines thinking** without expecting to be **contradicted**.

The Turing Test (Stanford Encyclopedia of Philosophy)

plato.stanford.edu/entries/turing-test/ ▾

Apr 9, 2003 - According to Turing, the question whether **machines can think** is itself to speak of **machines thinking** without expecting to be **contradicted**.

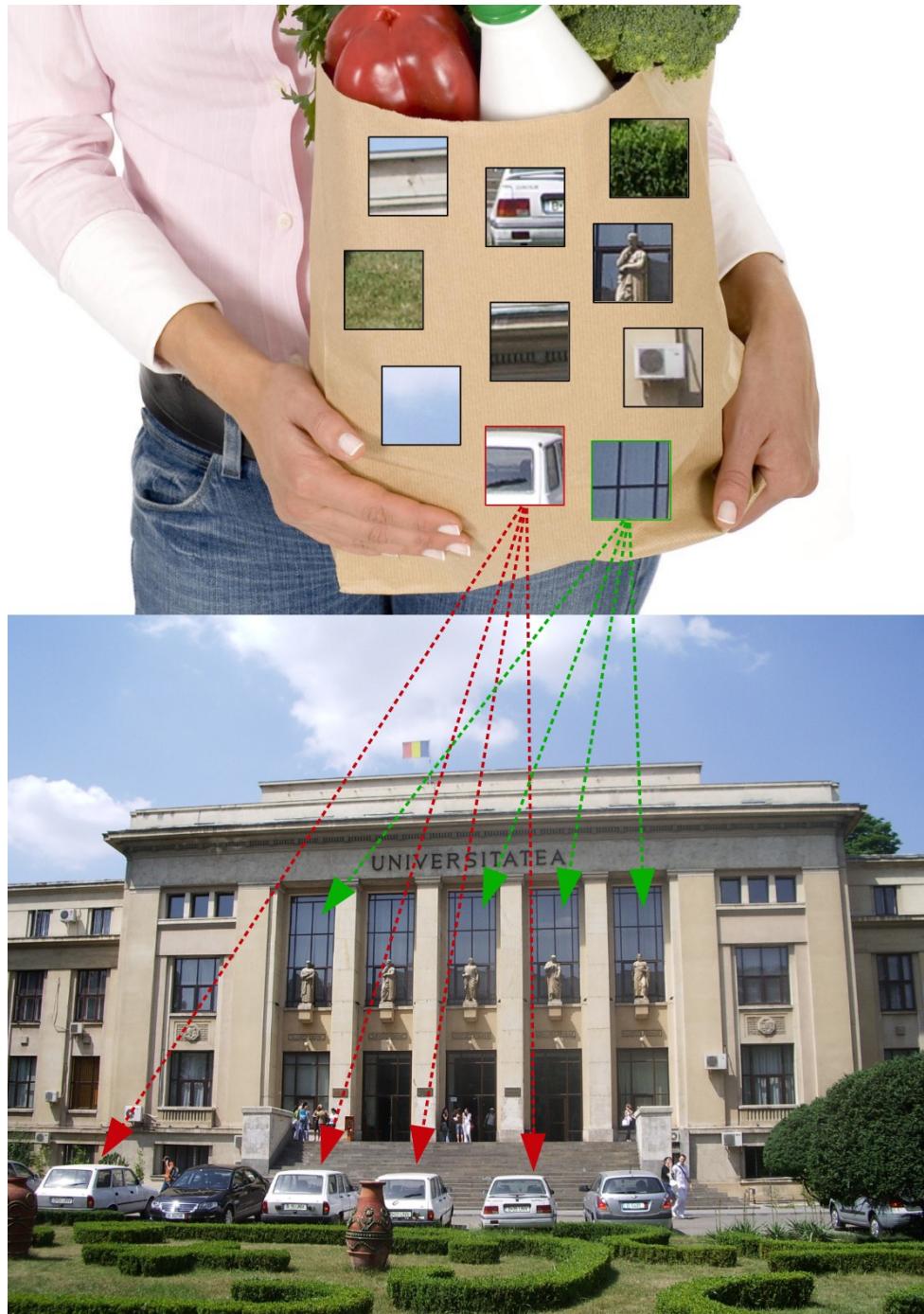
Turing: a natural philosopher: part 11

www.turing.org.uk/publications/ex11.html ▾

The remaining questions concern the computable discrete state **machine** one will be able to speak of **machines thinking** without expecting to be **contradicted**.

De la cuvinte la cuvinte vizuale

- Dacă am trata imaginile în mod asemănător, vom putea construi soluții performante pentru recunoaștere obiectelor
- Mai întâi trebuie să redefinim o serie de concepte:
vocabular – un set predefinit de cuvinte
cuvânt – un element distinct și cu înțeles al vorbirii
- Cuvintele sunt puse împreună pentru a genera diverse propoziții cu înțeles
- Problema se reduce la găsirea unor descriptori locali care împreună generează diverse imagini



Obiect

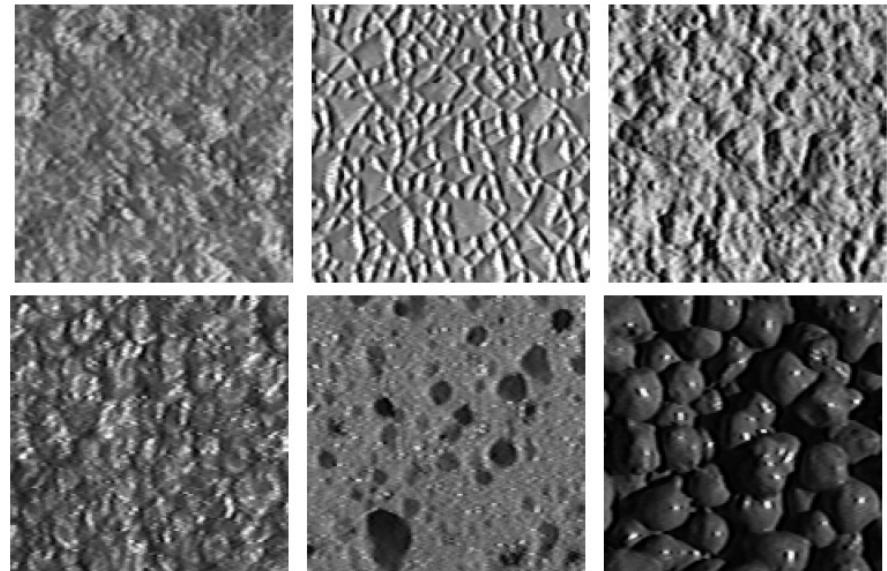
Bag of visual words

(sac de cuvinte vizuale)



Modele de tip bag of words: reprezentarea texturii

- Idee explorată pentru prima oară în reprezentarea materialelor și a texturii
- *Textoni* = centri clusterilor obținuți prin clusterizarea răspunsurile filtrelor aplicate unor imagini
- Reprezintă/descrie materiale și texturi pe baza distribuției textonilor (elemente prototip)



Recunoașterea texturii



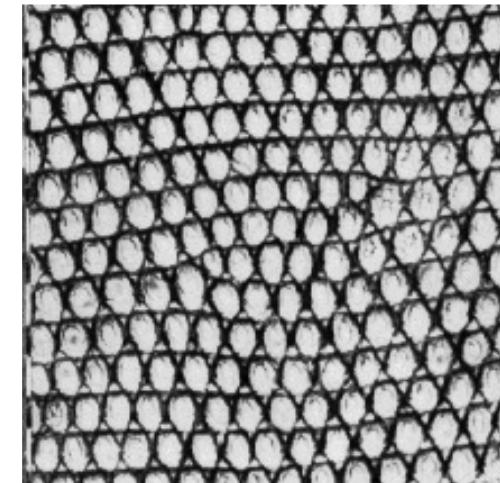
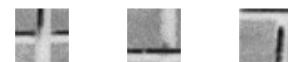
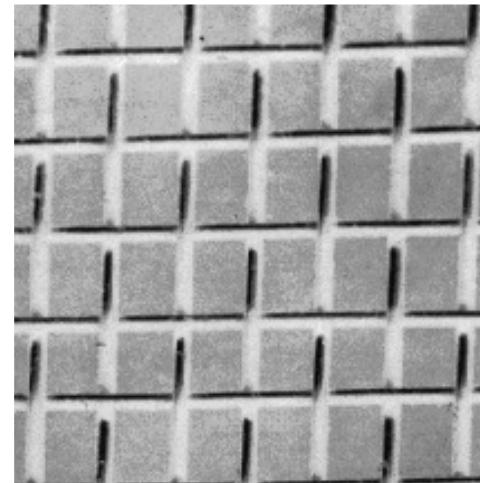
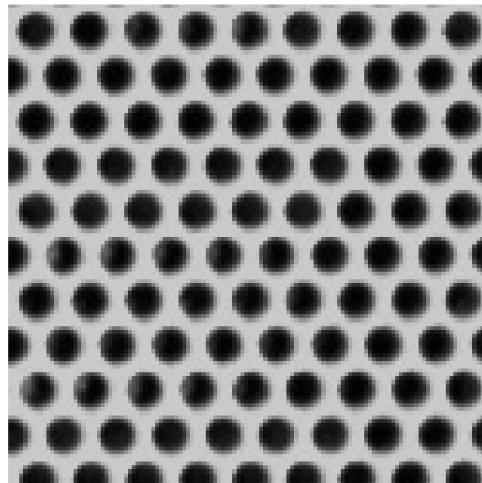
regulate

→ stocastice

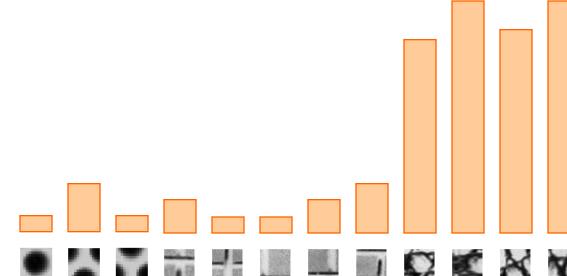
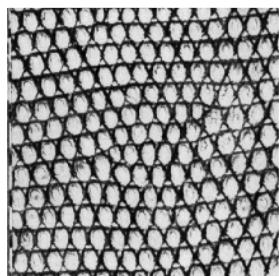
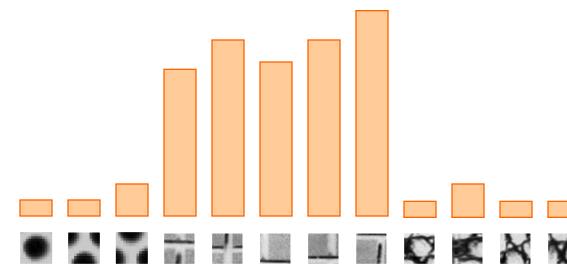
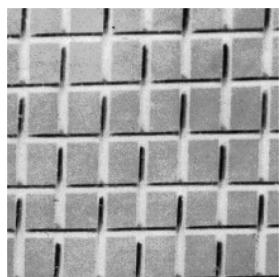
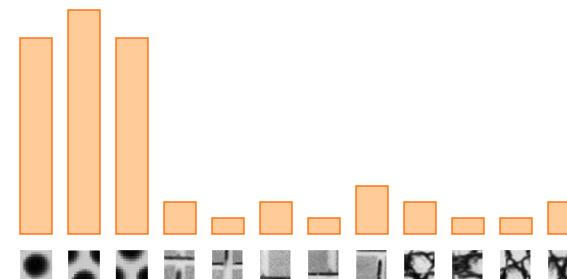
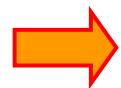
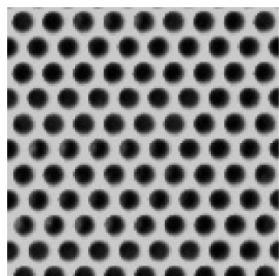
Exemple de texturi (Wikipedia)

Recunoașterea texturii

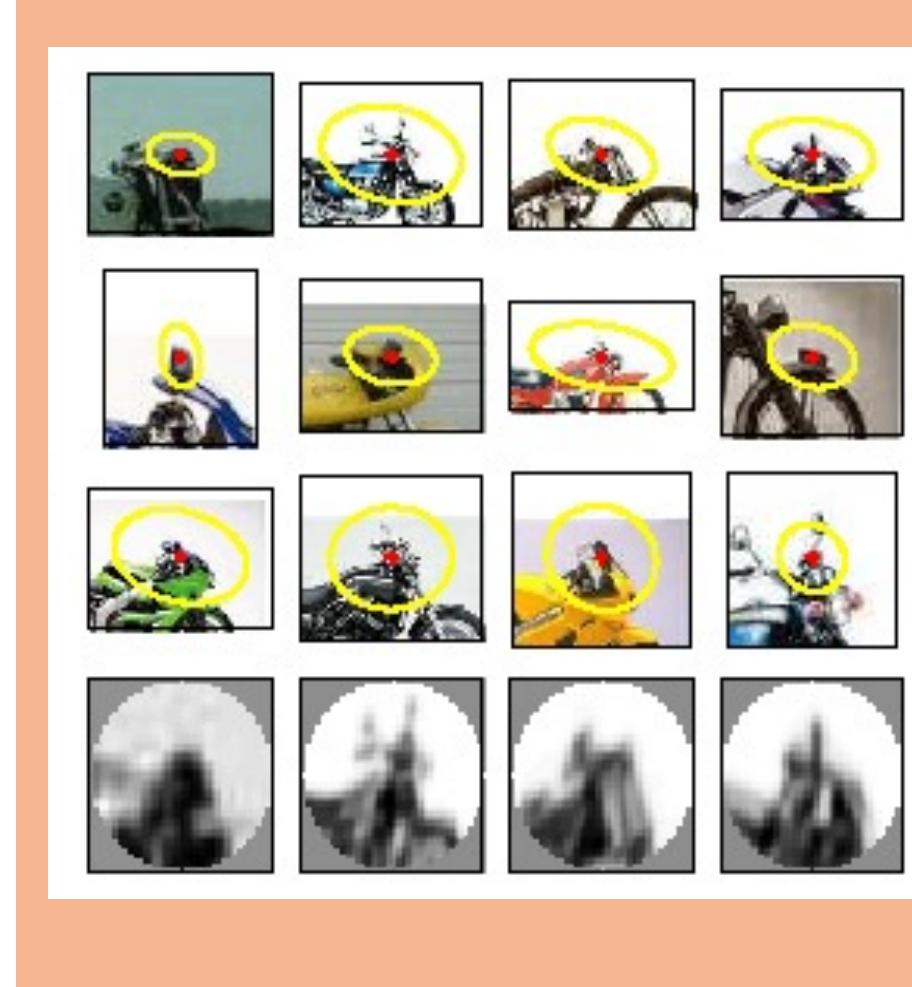
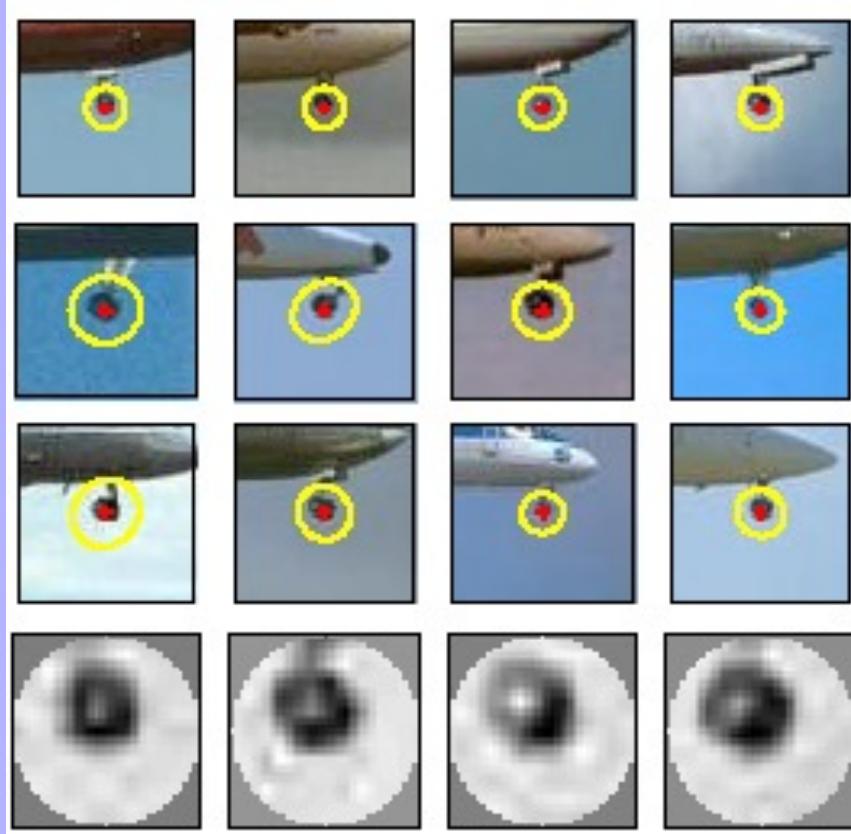
- Textura este caracterizată prin repetiția elementelor de bază, *textoni*
- Reprezentăm texturi prin histograme de textoni (numărăm tipul/câți textoni apar în fiecare imagine printr-o histogramă normalizată)



Recunoașterea texturii

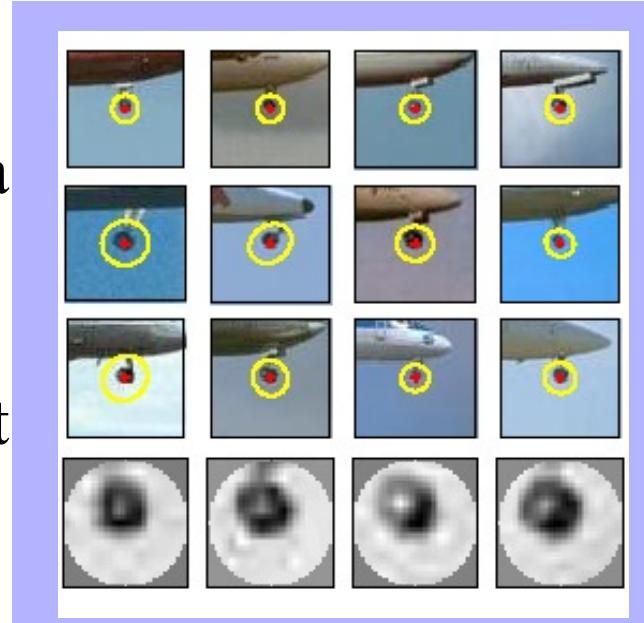


Visual words = cuvinte vizuale



Visual words = cuvinte vizuale

- Descriem cu ajutorul lor înfățisarea obiectelor
- Înfățisarea obiectelor variază foarte mult chiar și pentru aceeași clasă de obiecte
- Înfățisarea locală a părților componente variază mai puțin
- **Ideal: cuvânt vizual = parte a unui obiect**



Modelul ‘bag of visual words’

1. Multime de imagini

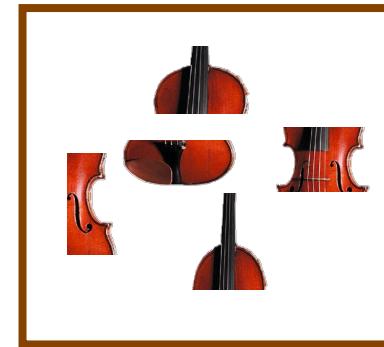
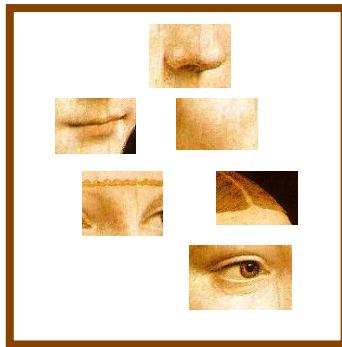


Modelul ‘bag of visual words’

1. Multime de imagini



2. Extragem caracteristici din fiecare imagine



Modelul ‘bag of visual words’

1. Multime de imagini



2. Extragem caracteristici din fiecare imagine
3. Învățăm un vocabular vizual = dicționar



Modelul ‘bag of visual words’

1. Multime de imagini



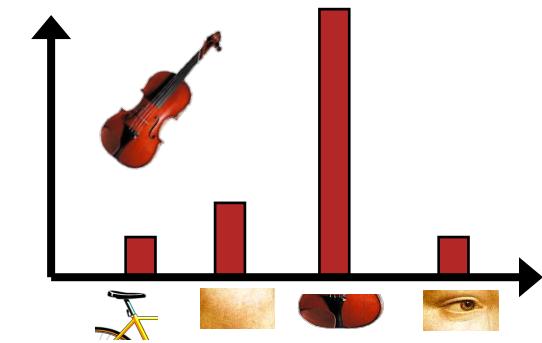
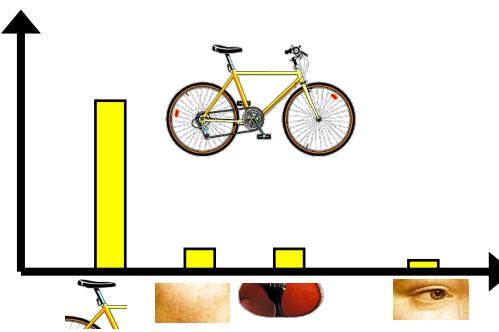
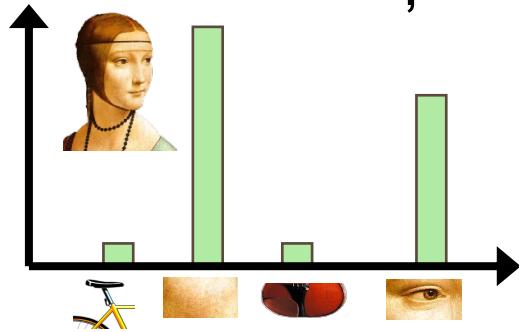
2. Extragem caracteristici din fiecare imagine
3. Învățăm un vocabular vizual = dicționar
4. Fiecare caracteristică este asignată “cuvântului vizual” cel mai apropiat din dicționar

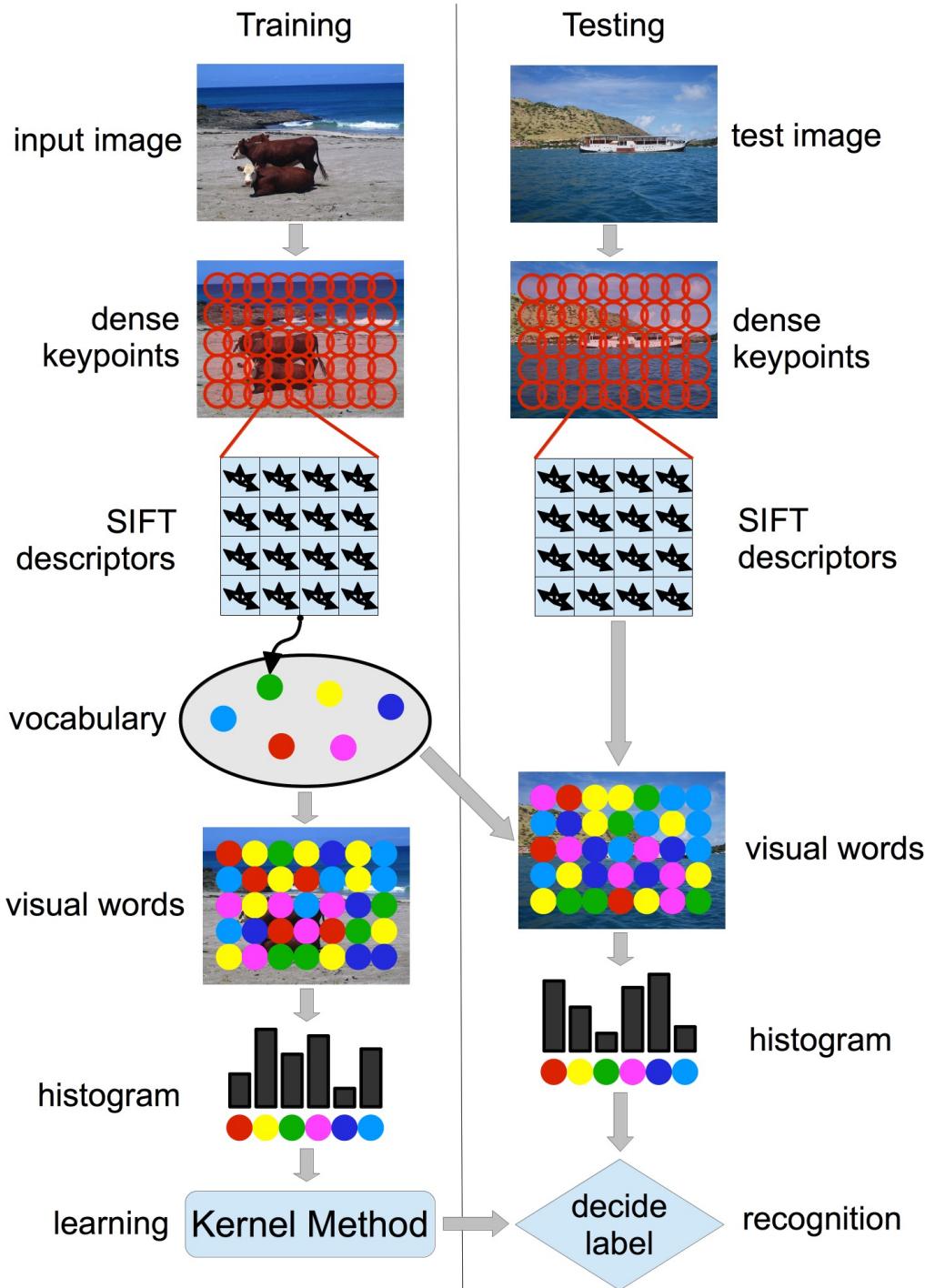
Modelul ‘bag of visual words’

1. Multime de imagini



2. Extragem caracteristici din fiecare imagine
3. Învățăm un vocabular vizual = dicționar
4. Fiecare caracteristică este asignată “cuvântului vizual” cel mai apropiat din dicționar
5. Reprezentăm imaginile pe baza unei histograme de frecvențe a “cuvintelor vizuale”

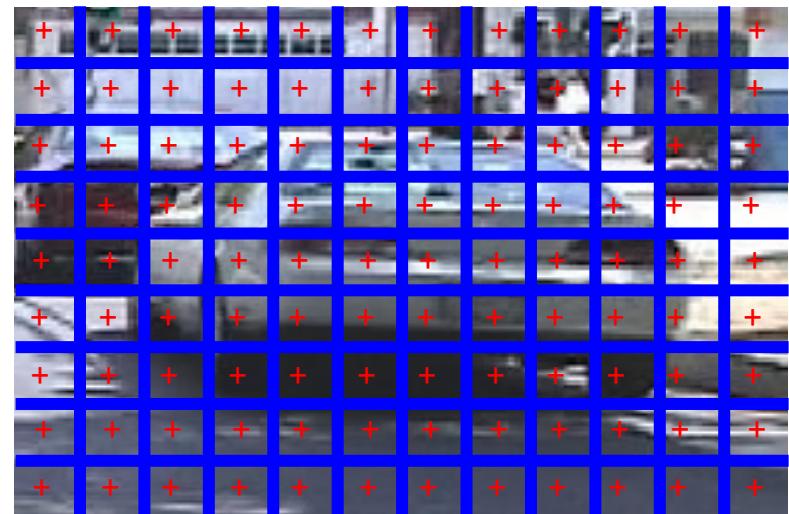




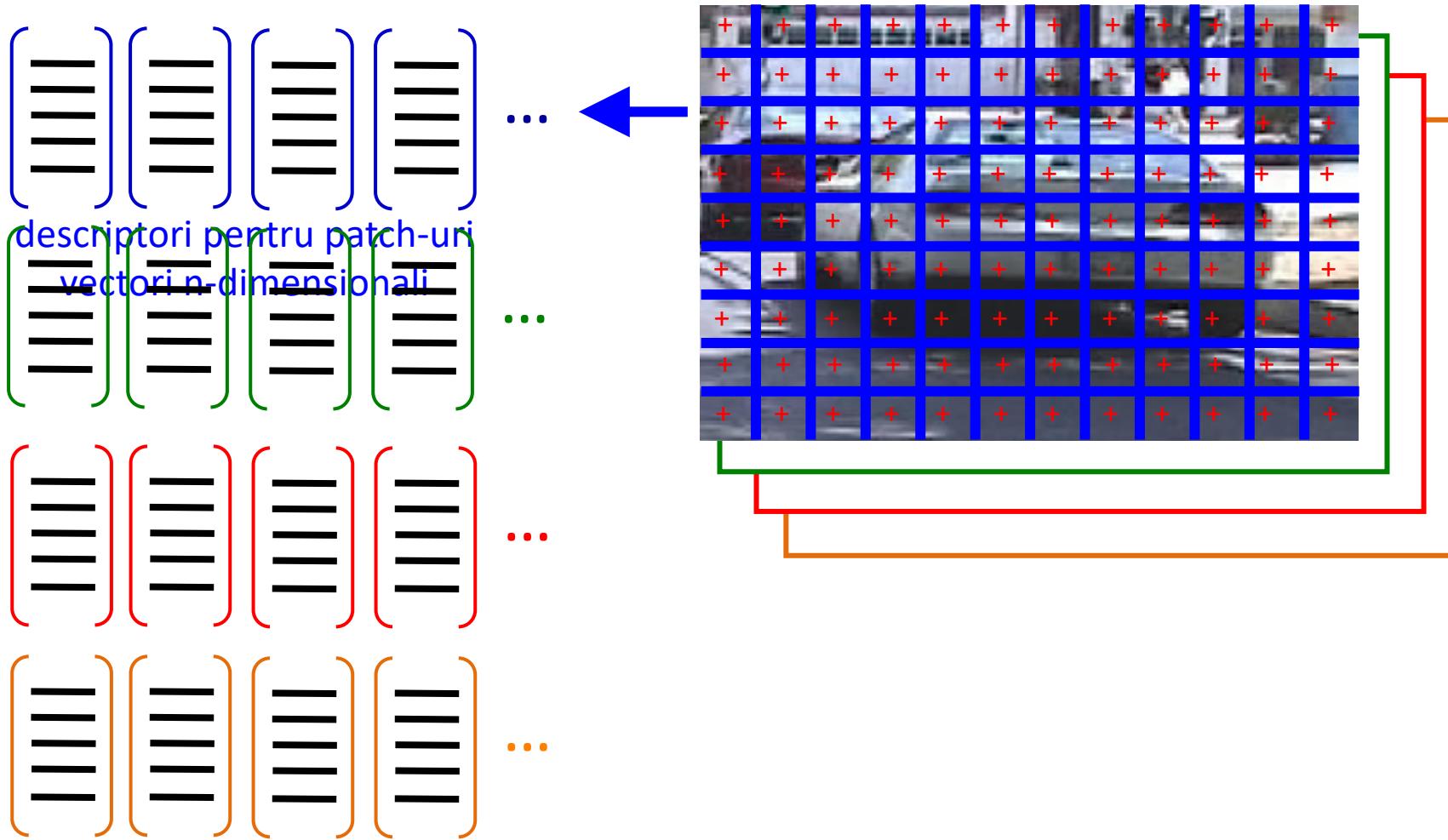
2. Extragerea de caracteristici

Puncte pe un caroaj, puncte ‘interesante’, puncte alese aleator în imagine

Descriem un **patch** (portiune mică din imagine) centrată pentru fiecare **punct ales**



2. Extragerea de caracteristici



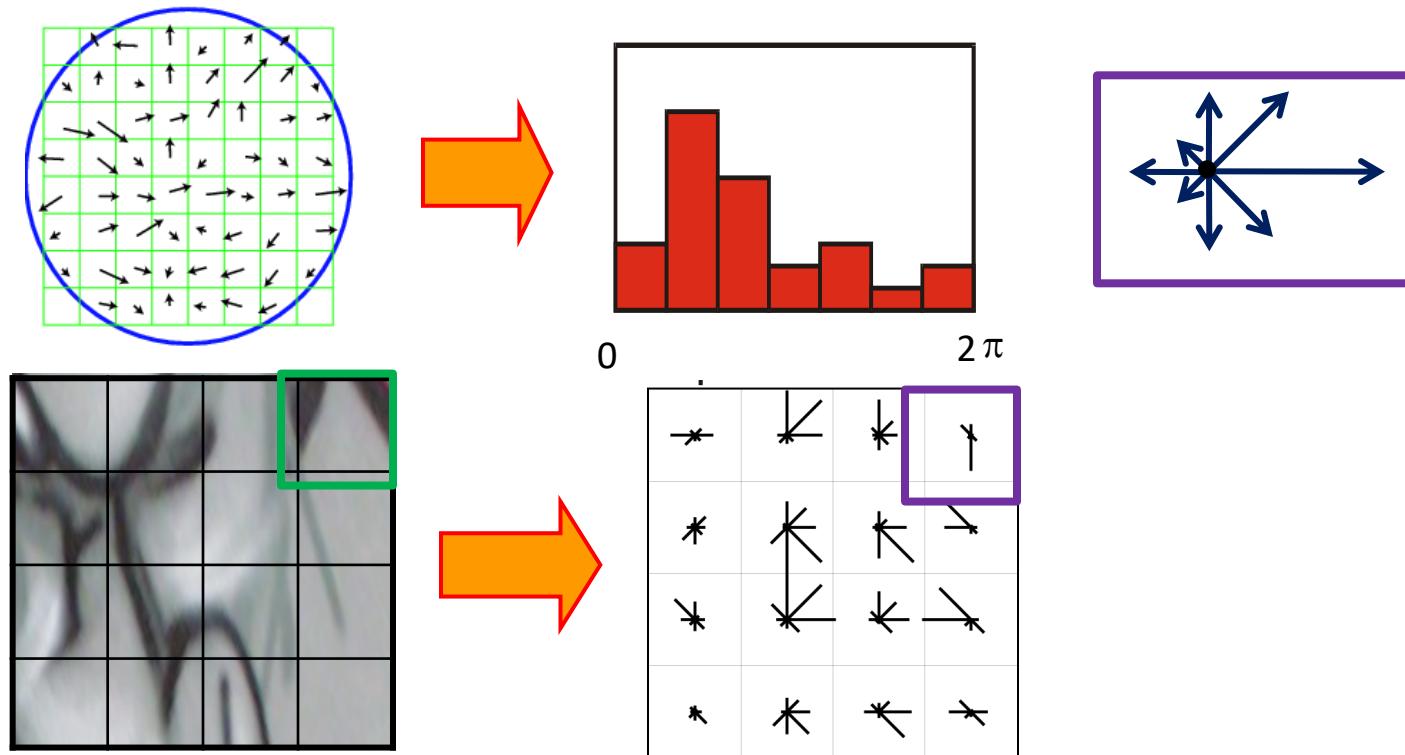
2. Extragerea de caracteristici

Trăsături utilizate des în practică:

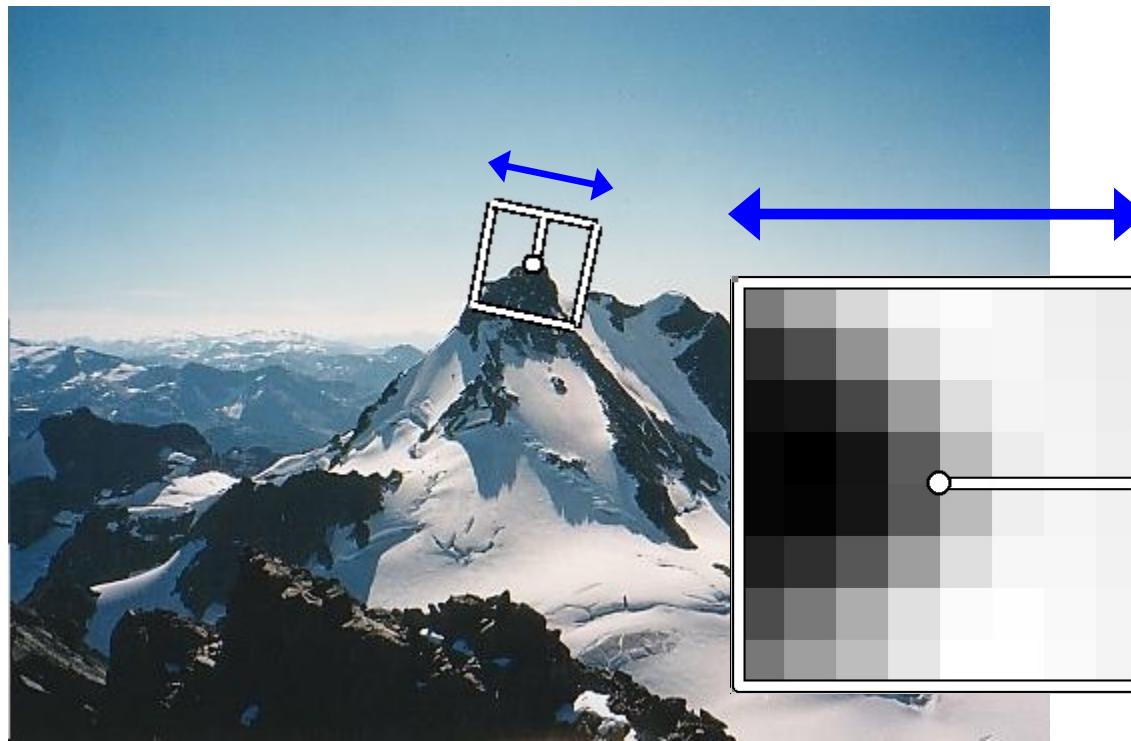
- SIFT – Scale Invariant Feature Transform
(în continuare)
- HOG – Histogram of Oriented Gradients
(în Cursul 8)

Descriptorul SIFT [Lowe 2004]

- împarte regiunea în 4×4 blocuri = 16 blocuri
- calculează pentru fiecare bloc o histogramă de orientări ale pixelilor (intervalul $0-360^0$ împărțit în 8 intervale)
- descriptorul SIFT = 16 blocuri = 16 histograme \times 8 valori = dimensiune 128



Descriptor invariant la rotație



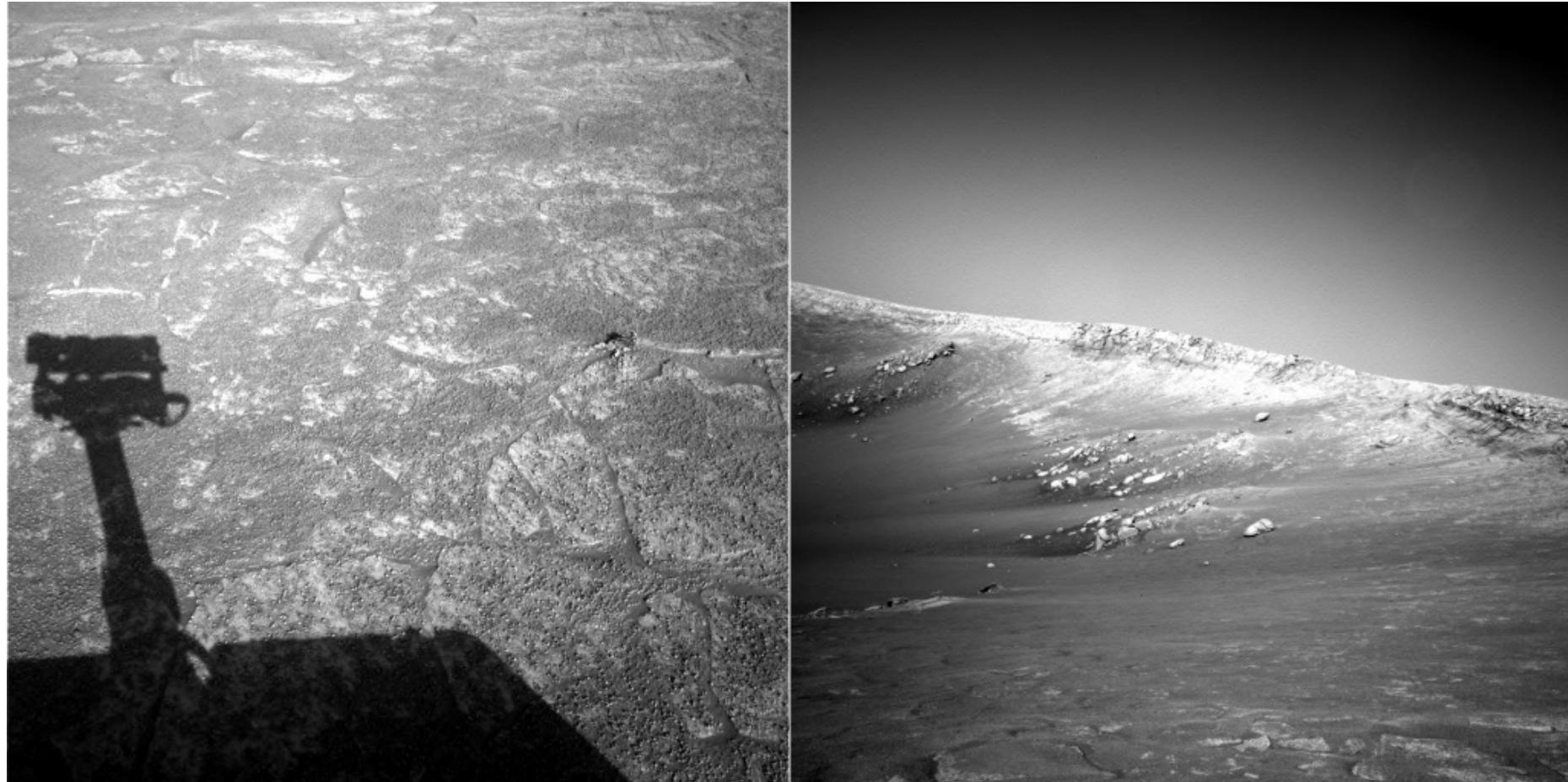
- rotim patch-ul (regiunea) pe baza orientării dominante
- obțin un patch cu orientare canonică

Descriptorul SIFT [Lowe 2004]

- foarte robust pentru matching (găsire de corespondențe)
- robust la schimbări mari ale unghiului din care a fost luată imaginea
- robust la schimbări mari în iluminarea scenei
 - chiar și condiții de zi vs condiții de noapte
- foarte rapid – poate rula în timp real
- cod disponibil online (OpenCV)

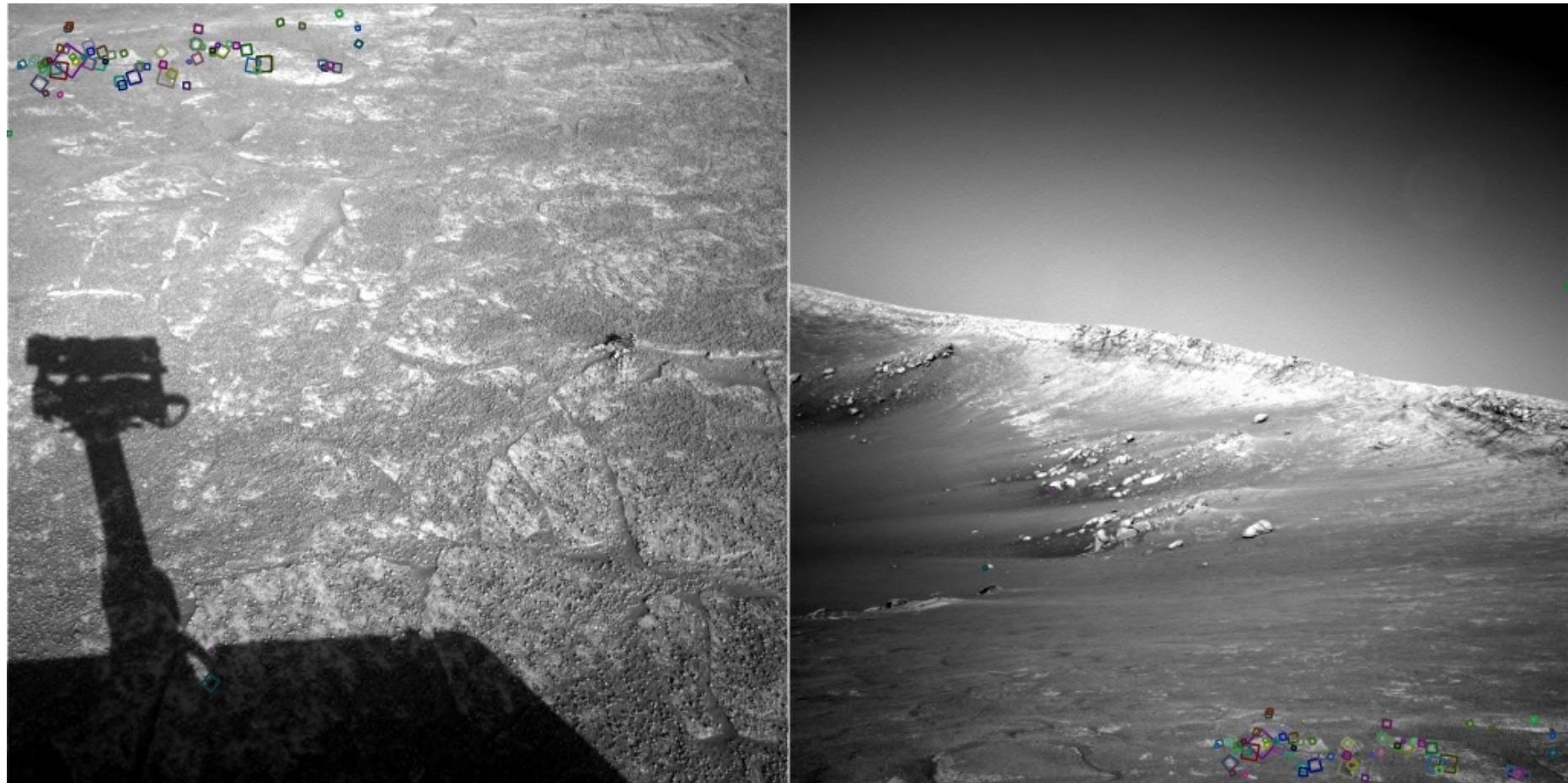


Descriptorul SIFT [Lowe 2004]



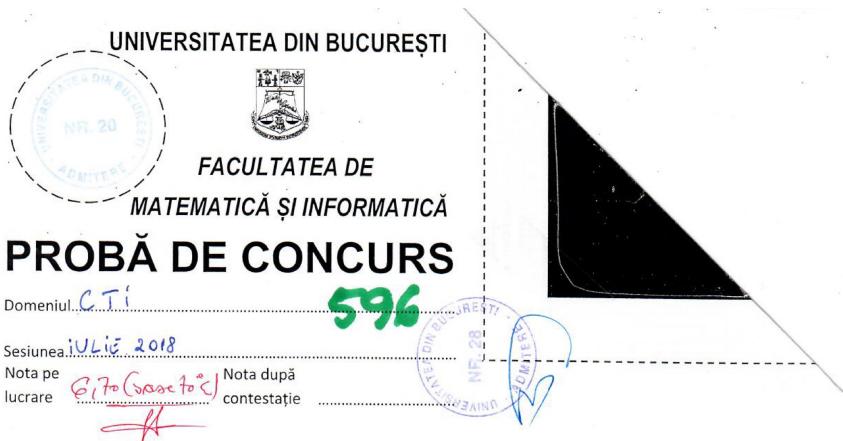
NASA Mars Rover images

Descriptorul SIFT [Lowe 2004]



NASA Mars Rover images
with SIFT feature matches
Figure by Noah Snavely

Corectarea automată a testelor grilă



TEST GRILĂ

INFORMATICA

FIZICĂ

MATEMATICĂ

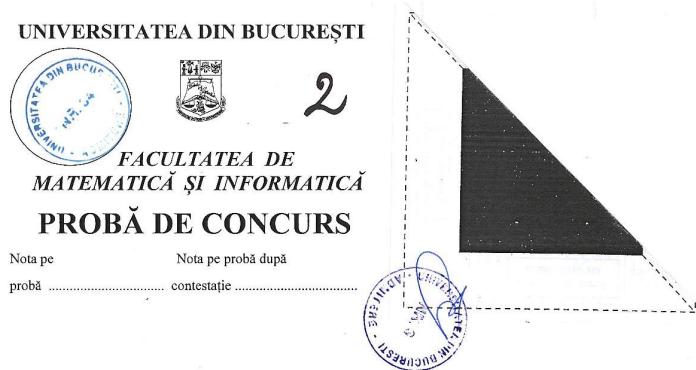
Număr întrebare	Răspuns	A	B	C	D
1		X			
2		X			
3		X			
4		X			
5			X		
6		X			
7		X			
8				X	
9			X		
10		X			
11			X		
12			X		
13		X			
14				X	
15				X	

Număr întrebare	Răspuns	A	B	C	D
1				X	
2			X		
3		X			
4		X			
5			X		
6		X			
7				X	
8				X	
9			X		
10		X			
11		X			
12			X		
13			X		
14			X		
15				X	

NOTĂ : Se bifează X în căsuța corespunzătoare răspunsului corect.

image_2.txt	
1	3
2	B
3	A
4	B
5	C
6	A
7	B
8	D
9	C
10	A
11	C
12	C
13	A
14	D
15	D
16	C
17	B
18	A
19	A
20	B
21	A
22	D
23	D
24	C
25	A
26	A
27	B
28	B
29	B
30	C
R	19

Corectarea automată a testelor grilă

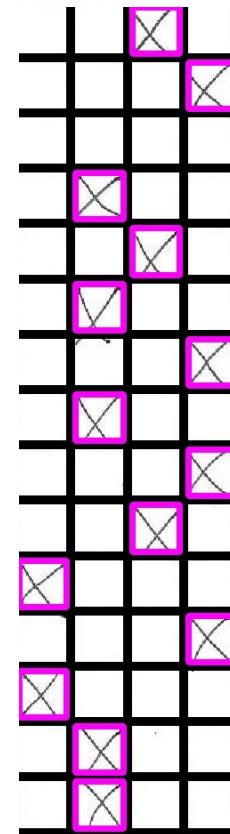


TEST GRILĂ

INFORMATICĂ 1 **FIZICĂ** _____
Număr varianță

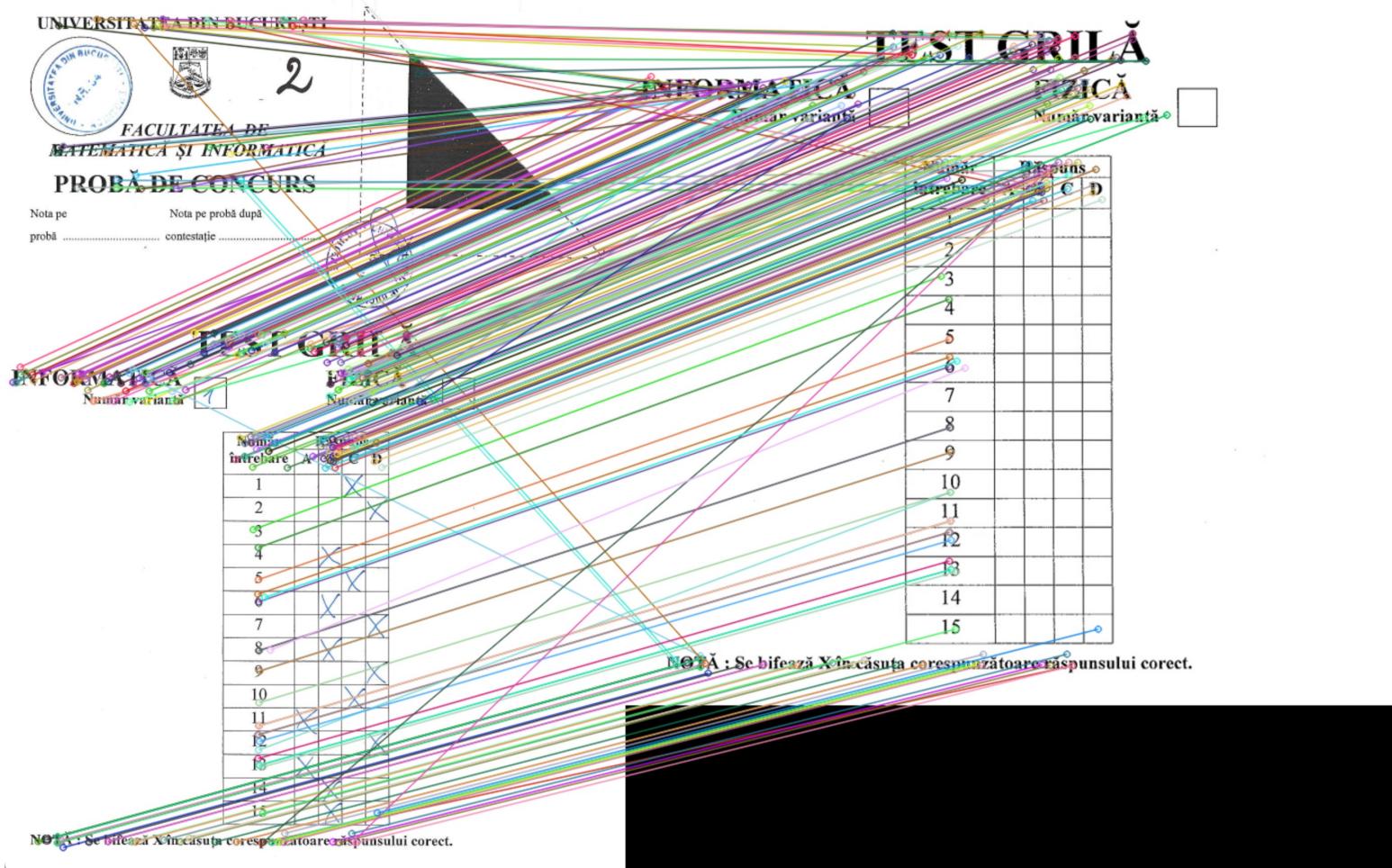
Număr întrebare	A	B	C	D
1			X	
2			X	
3				
4	X			
5		X		
6	X			
7		X		
8	X			
9			X	
10		X		
11	X			
12			X	
13	X			
14		X		
15		X		

NOTĂ : Se bifează X în căsuța corespunzătoare răspunsului corect.

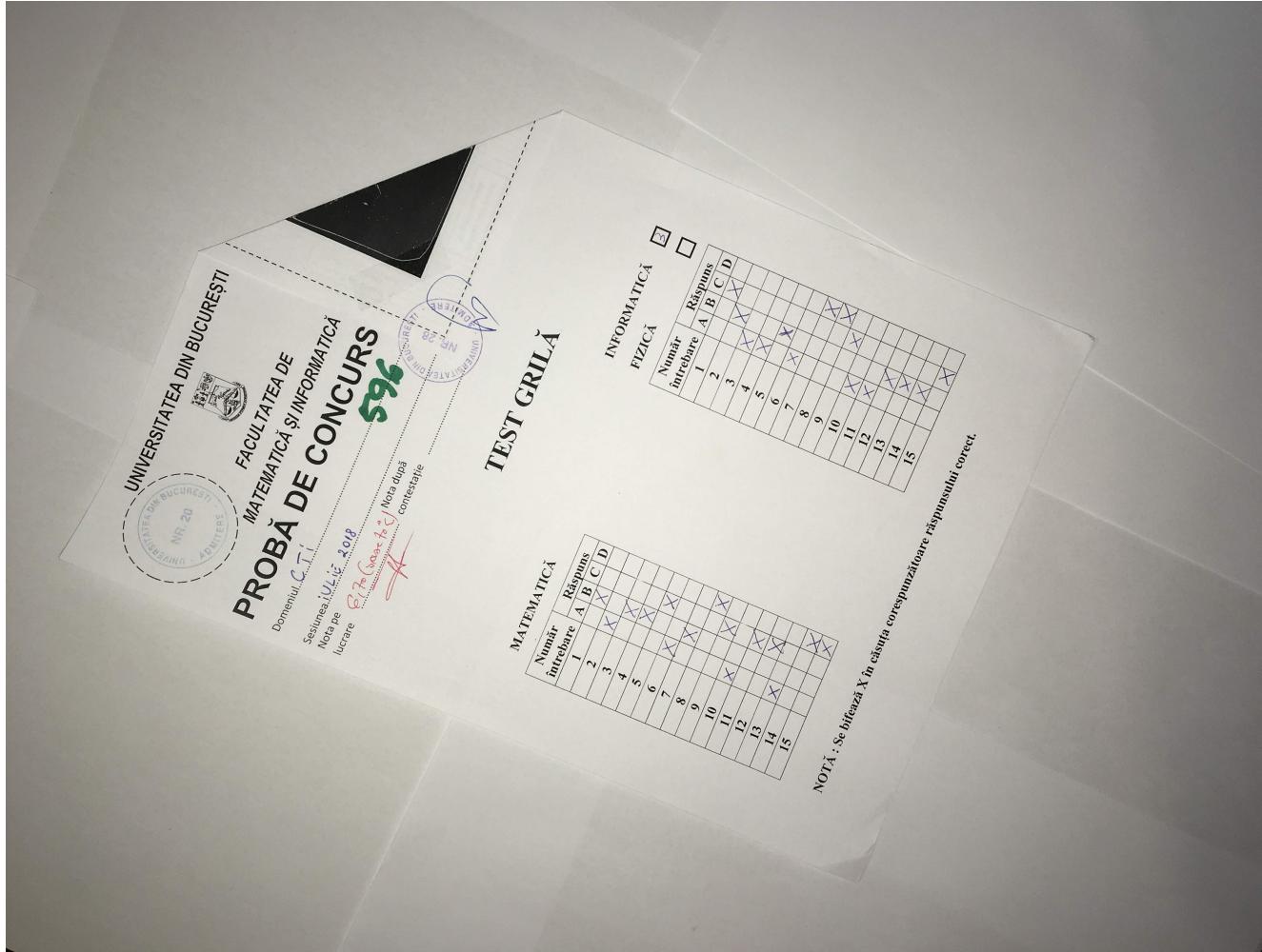


Numar Lucrare	Varianta	0X	1X	mX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Numar grile corecte	Nota
2	Informatica_1	1	14	0	C	D	B	C	B	D	B	D	C	A	D	A	B	B		2	2.2

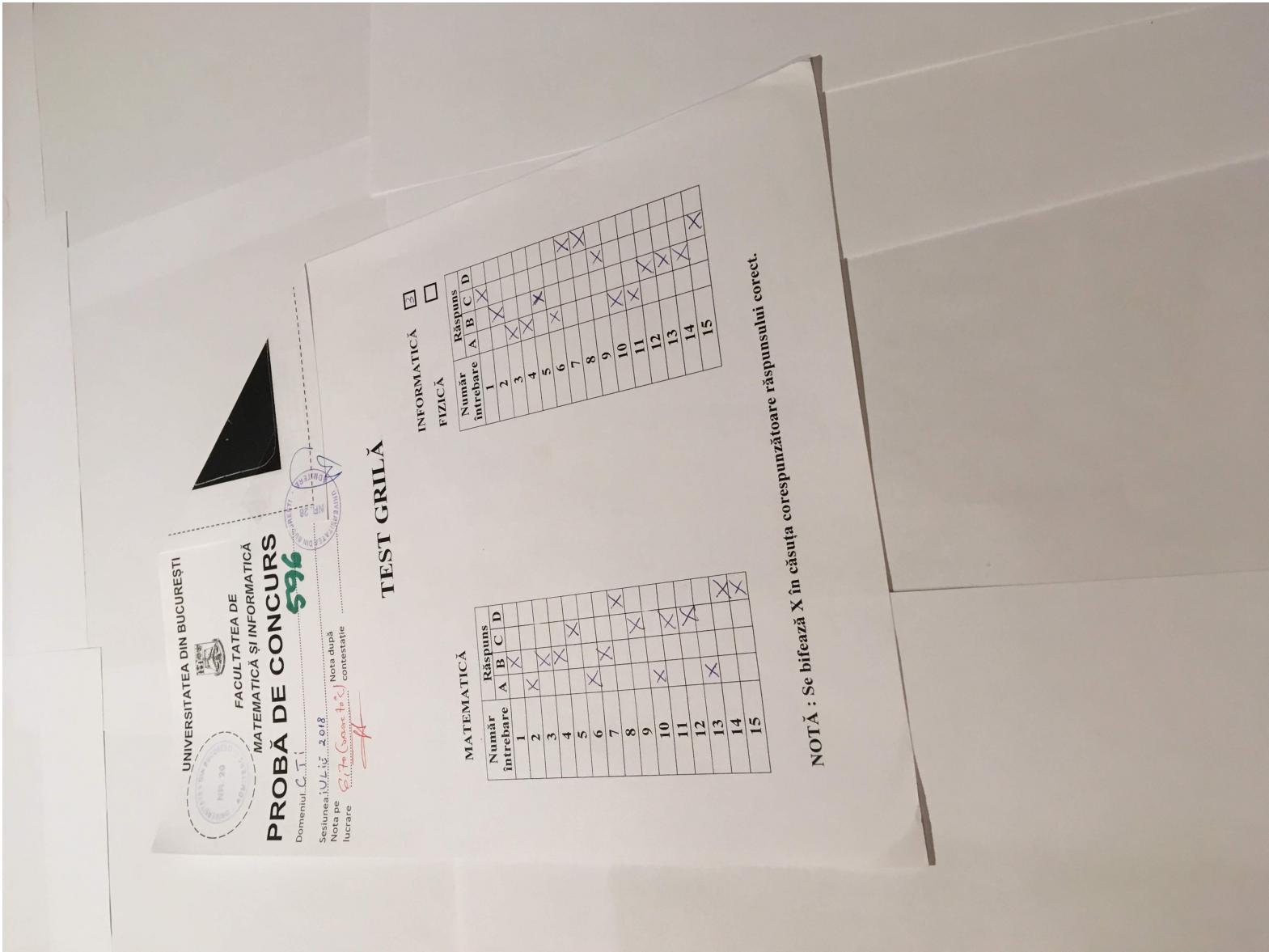
Corectarea automată a testelor grilă



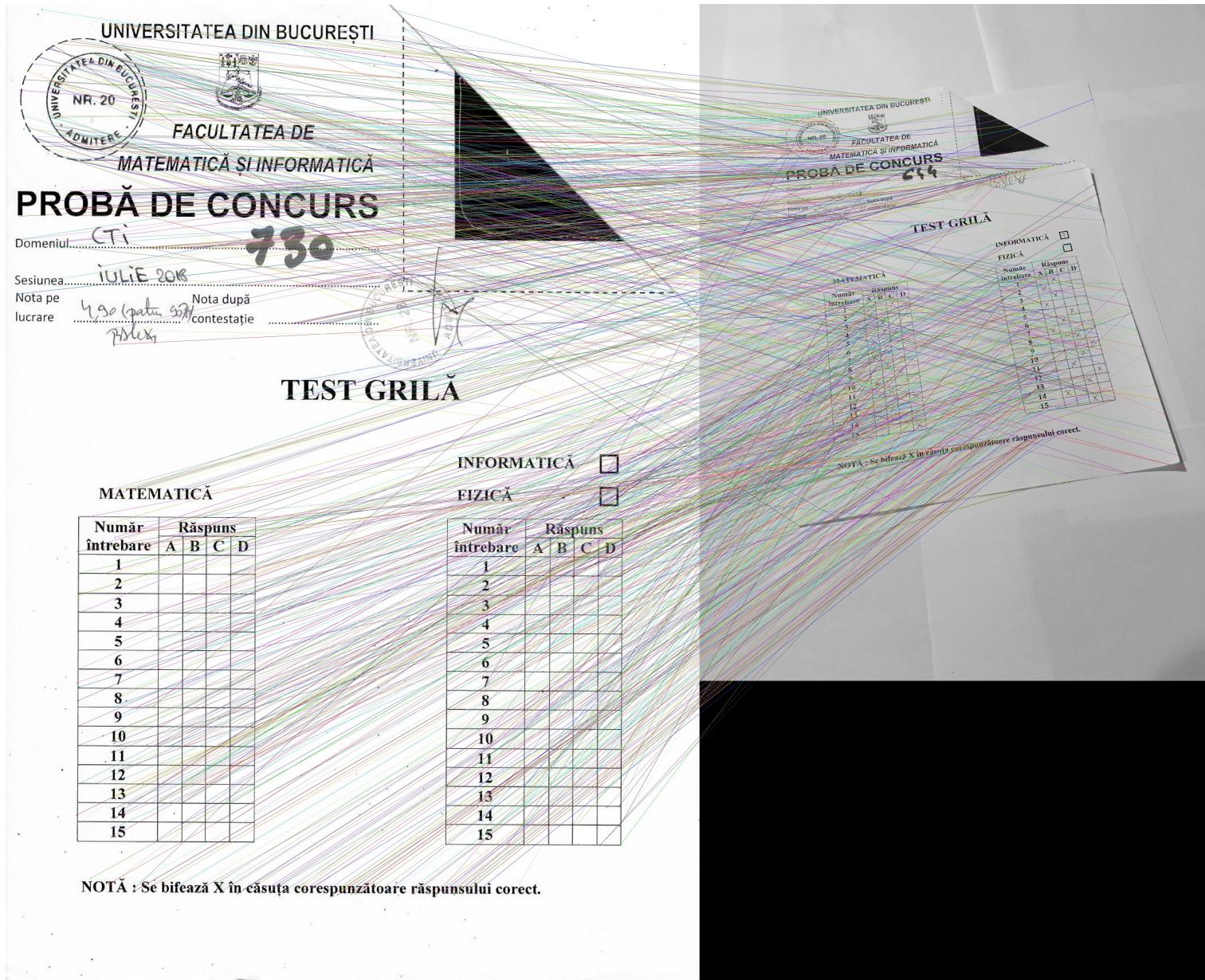
Rotație + scală



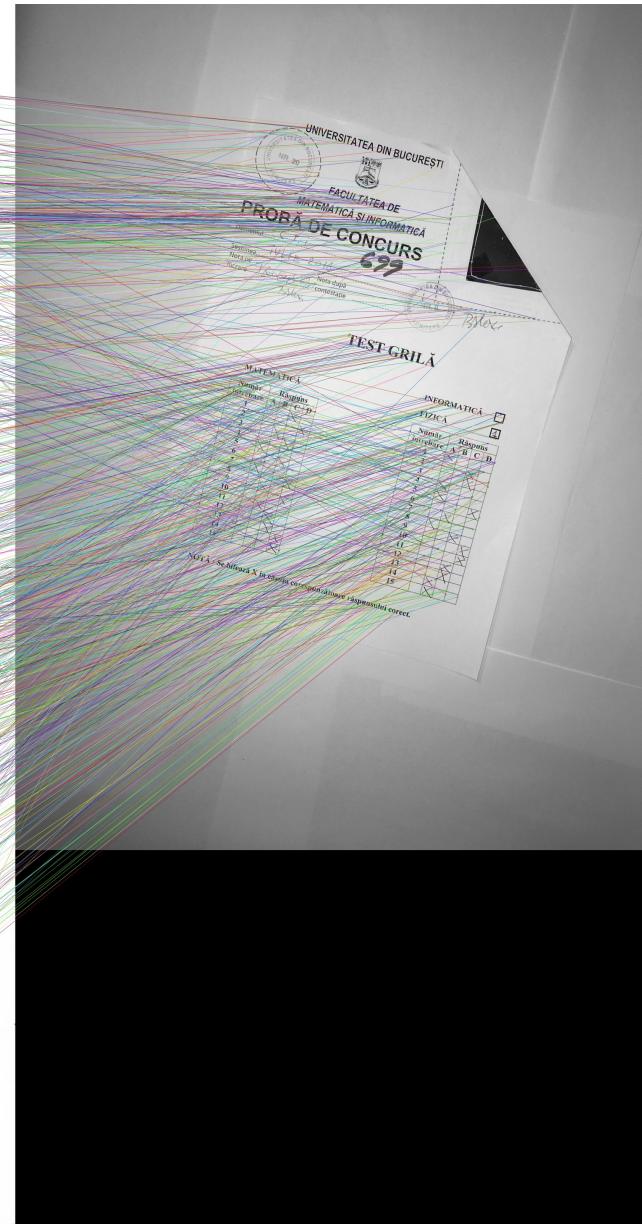
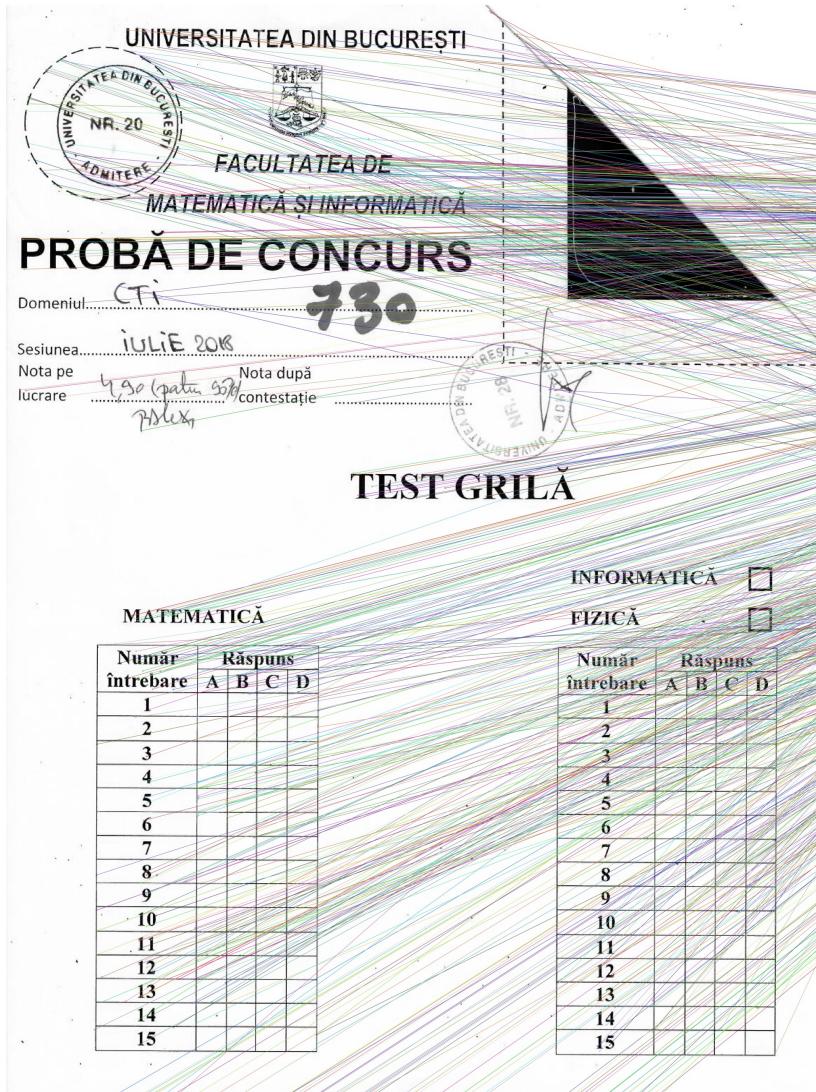
Perspectivă



Găsire de corespondențe cu SIFT



Găsire de corespondențe cu SIFT



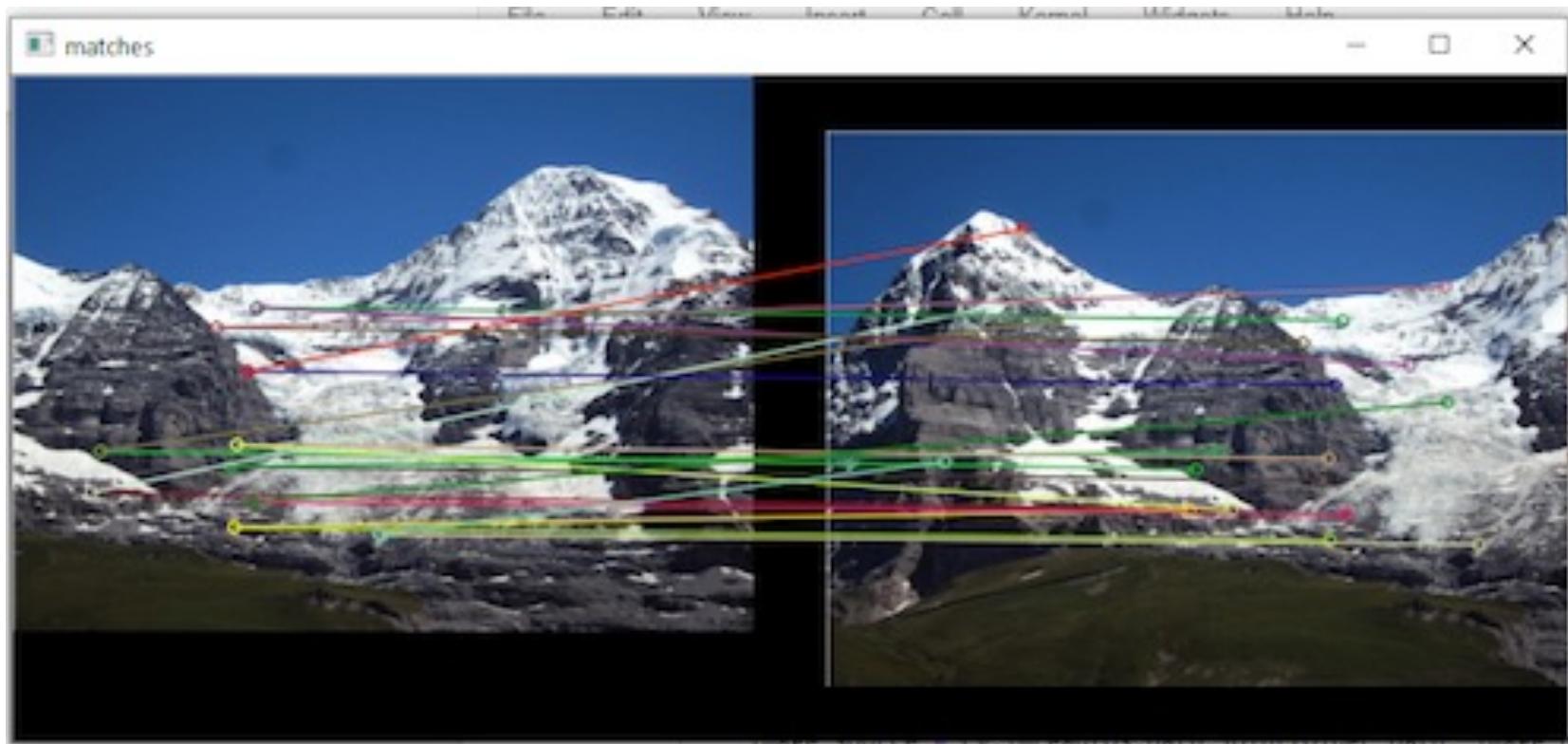
Realizarea de panorame



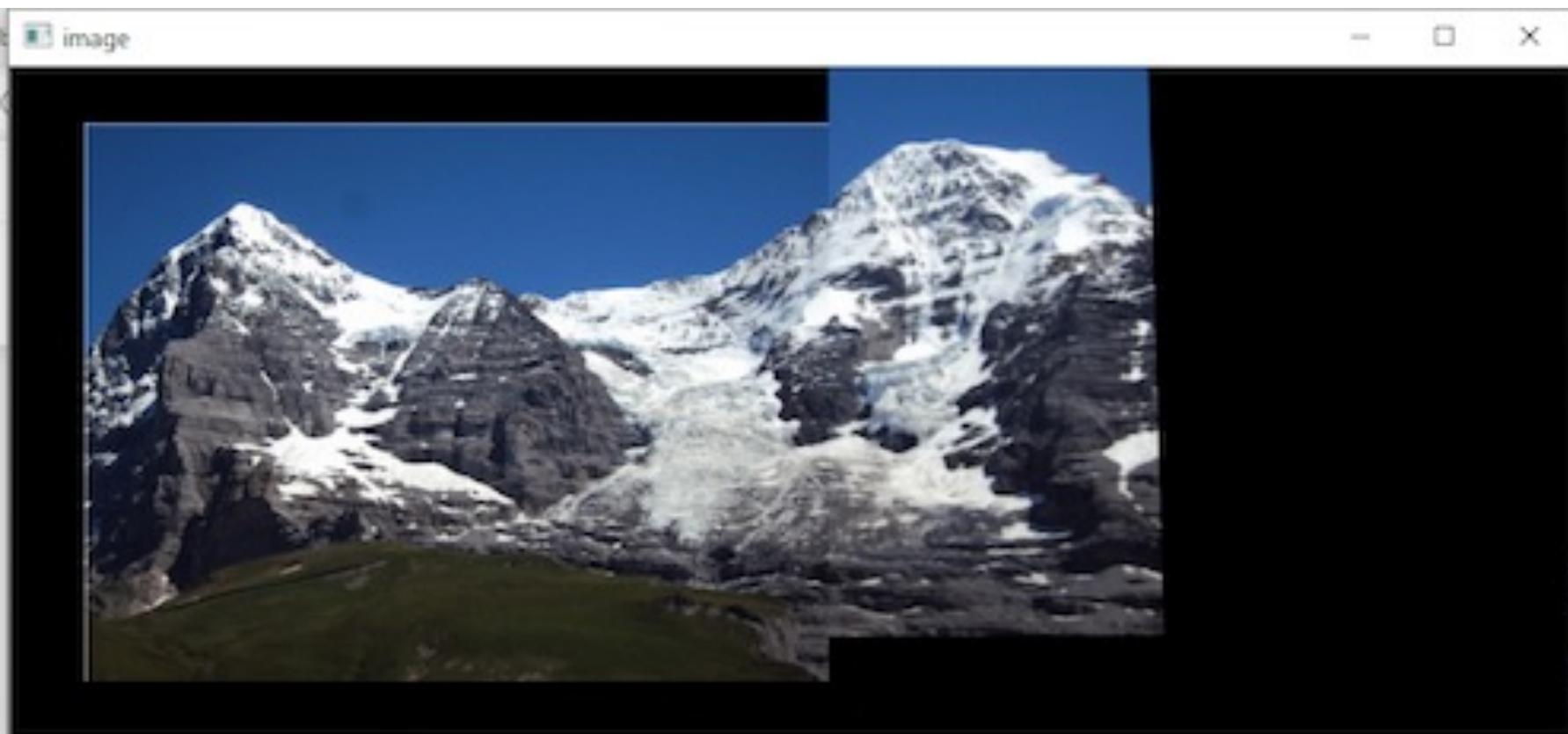
Realizarea de panorame



Realizarea de panorame



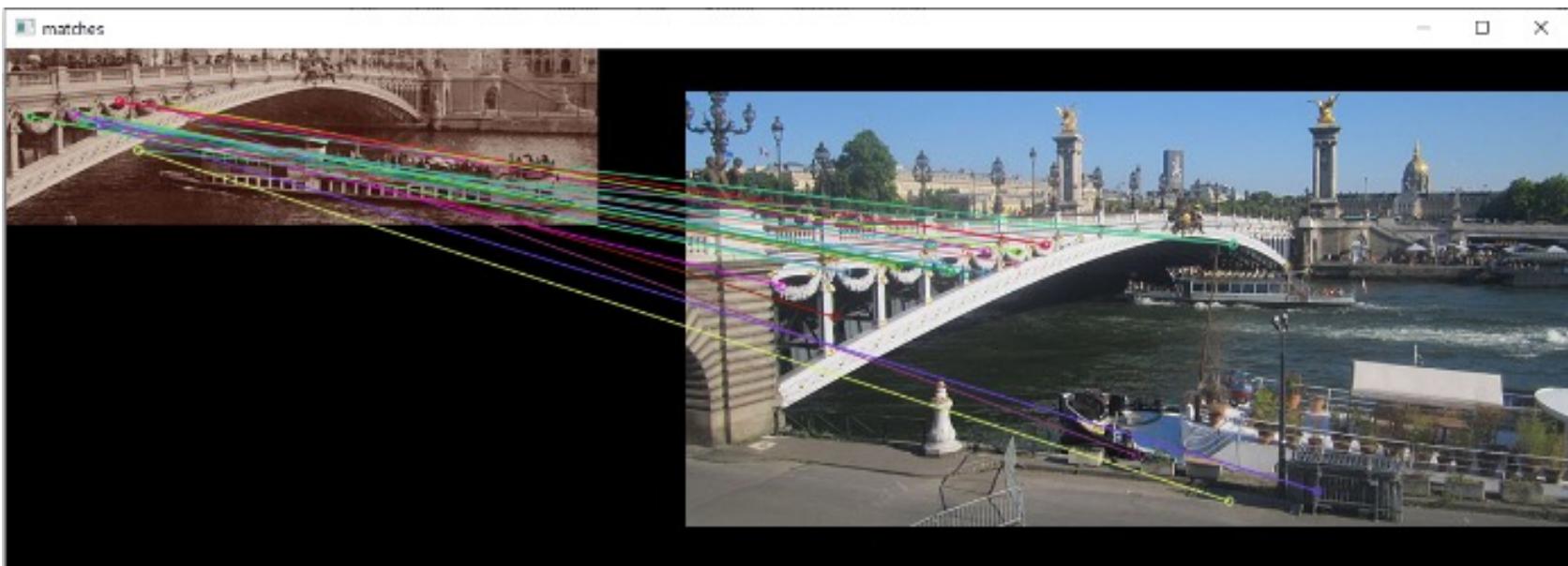
Realizarea de panorame



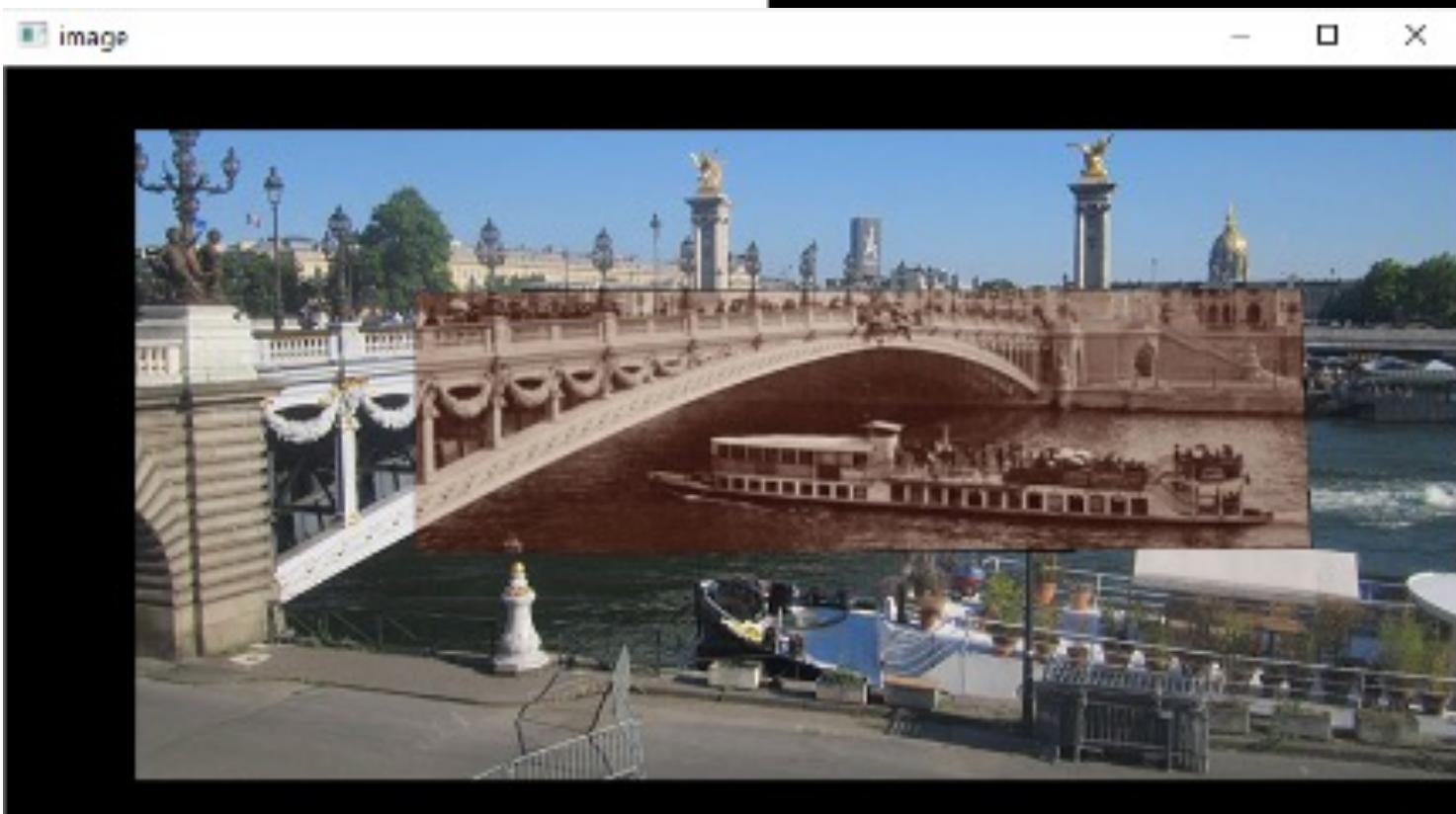
Realizarea de panorame



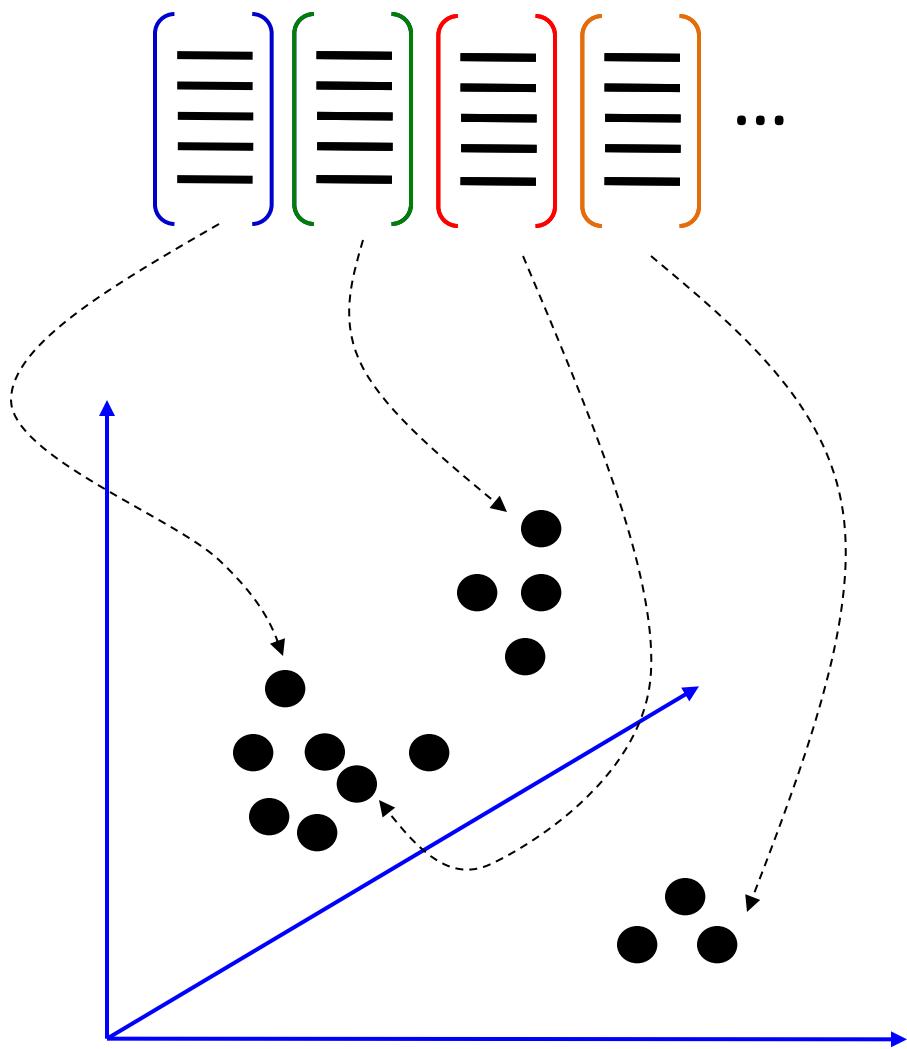
Then and now



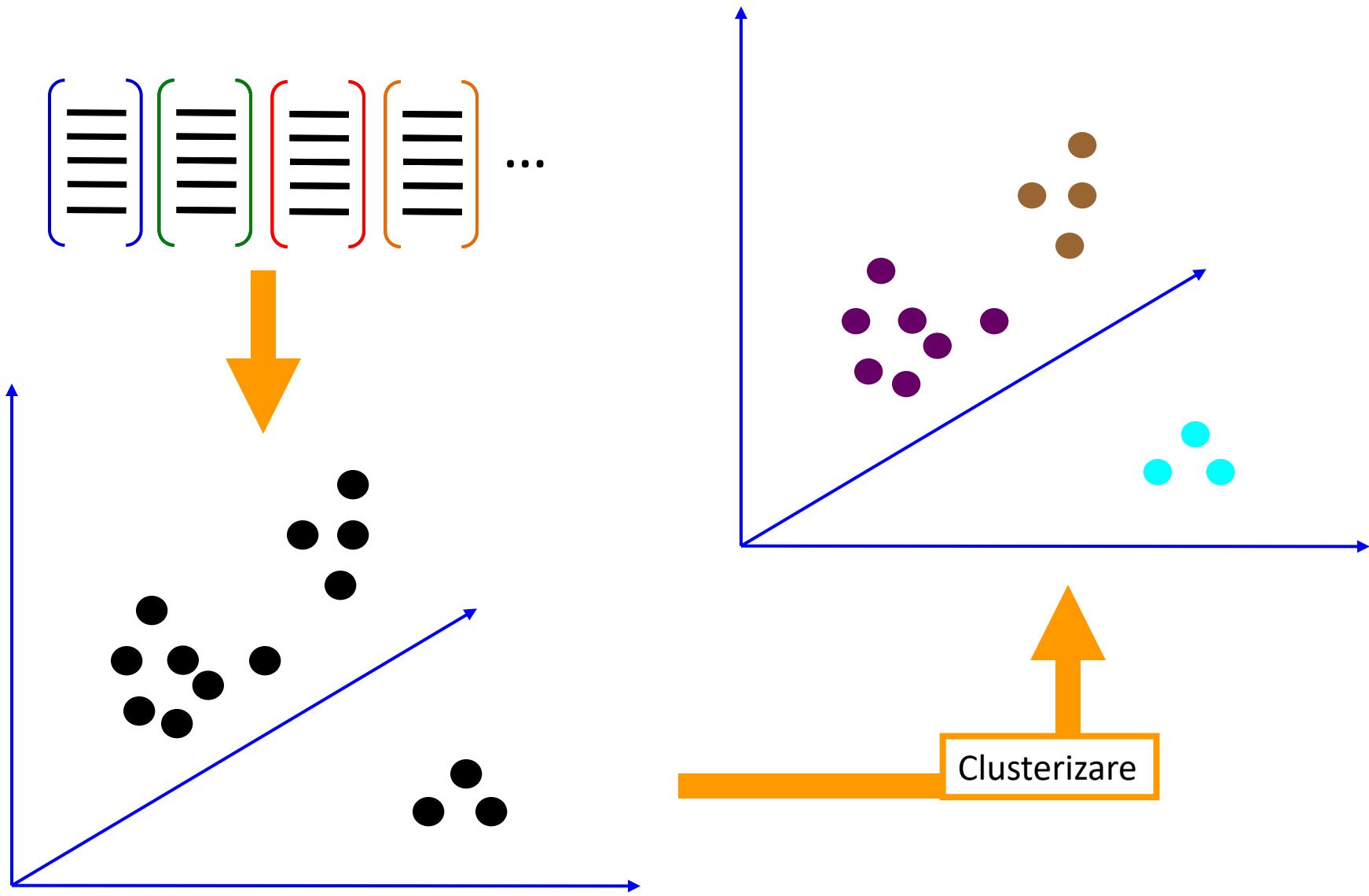
Then and now



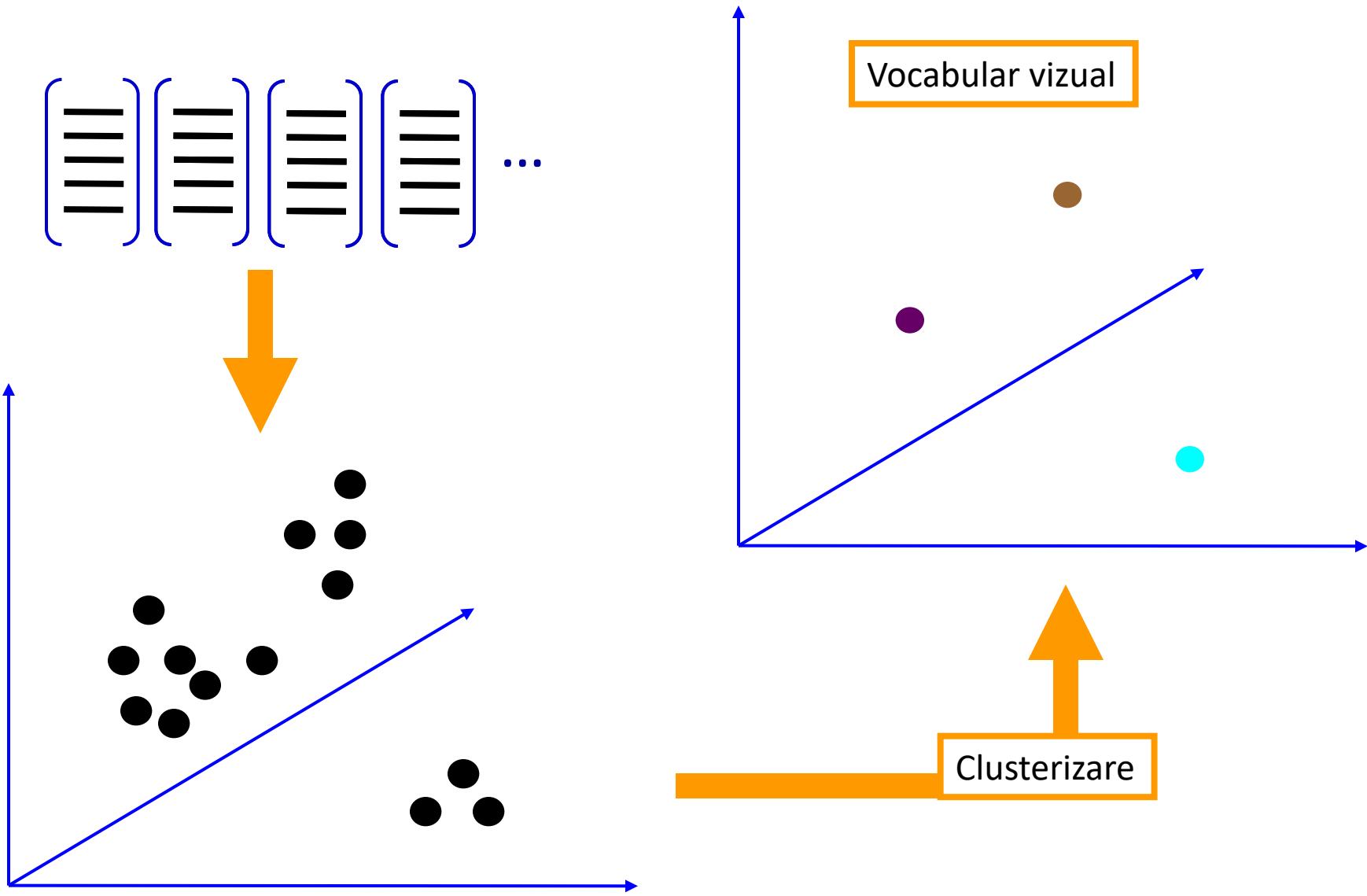
3. Învățarea vocabularului vizual



3. Învățarea vocabularului vizual



3. Învățarea vocabularului vizual



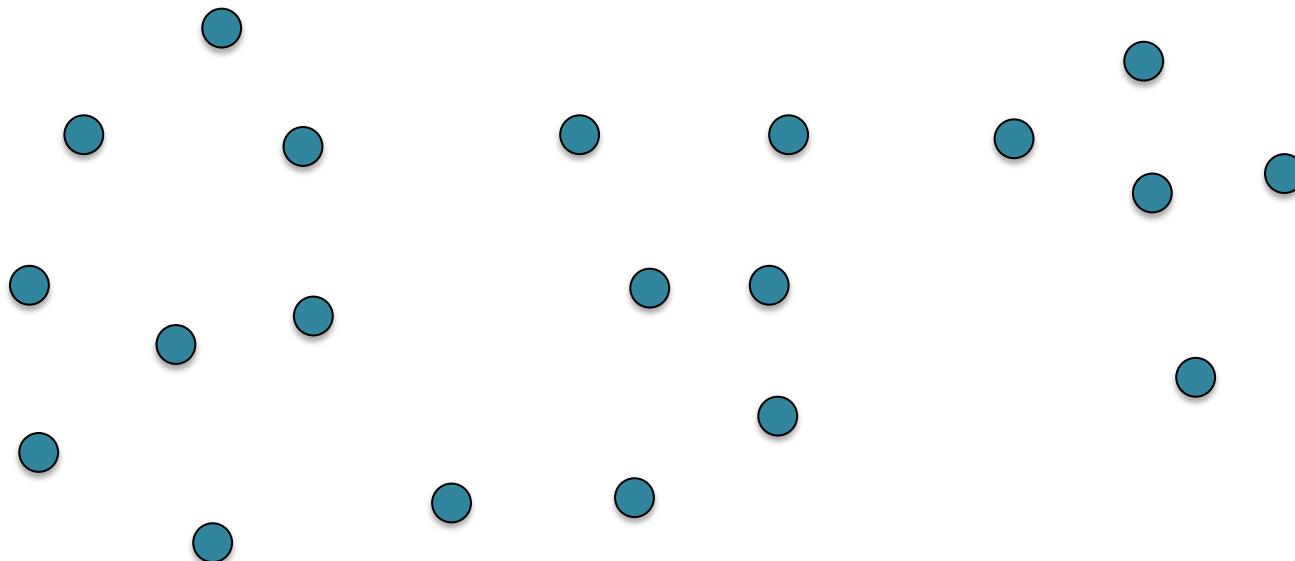
Clusterizarea K-means

- Vrem să minimizăm suma pătratelor distanțelor Euclidiene dintre caracteristicile \mathbf{x}_i și cei mai apropiati centri ai clusterilor \mathbf{m}_k

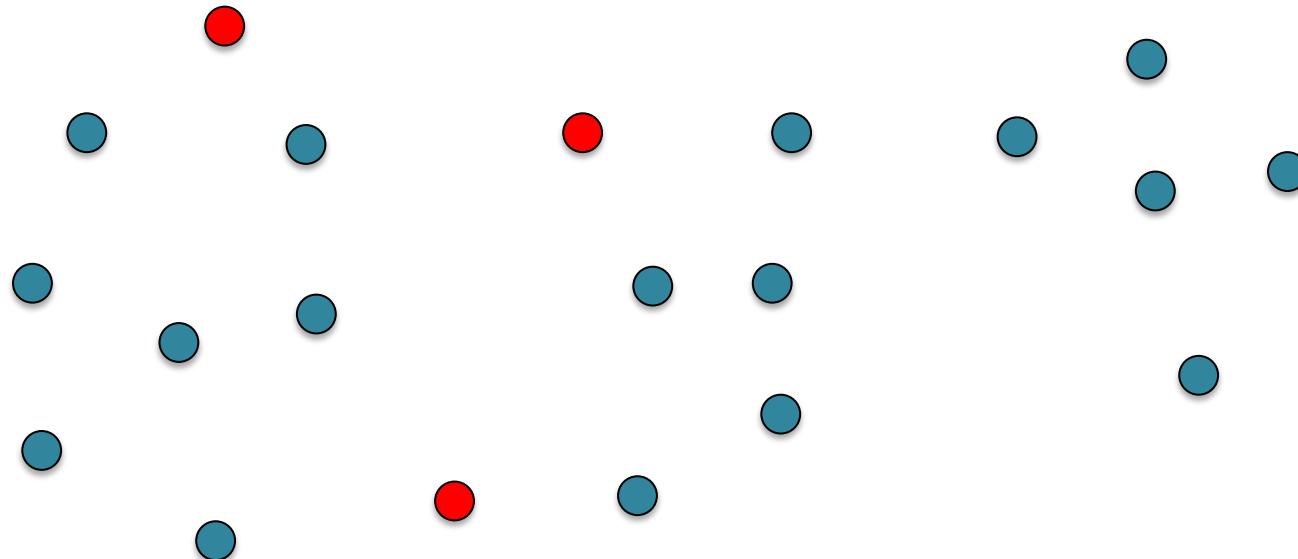
$$D(X, M) = \sum_{\text{clusterul } k} \sum_{\substack{\text{punctul } i \text{ în} \\ \text{clusterul } k}} (\mathbf{x}_i - \mathbf{m}_k)^2$$

- Algoritm:
 - 1) Inițializăm aleator cei K centri ai clusterelor
 - 2) Iterăm până când clusterele nu se modifică:
 - etichetăm fiecare caracteristică în funcție de cel mai apropiat centru al unui cluster
 - recalculăm centrul fiecărui cluster = media tuturor caracteristicilor cu aceeași etichetă

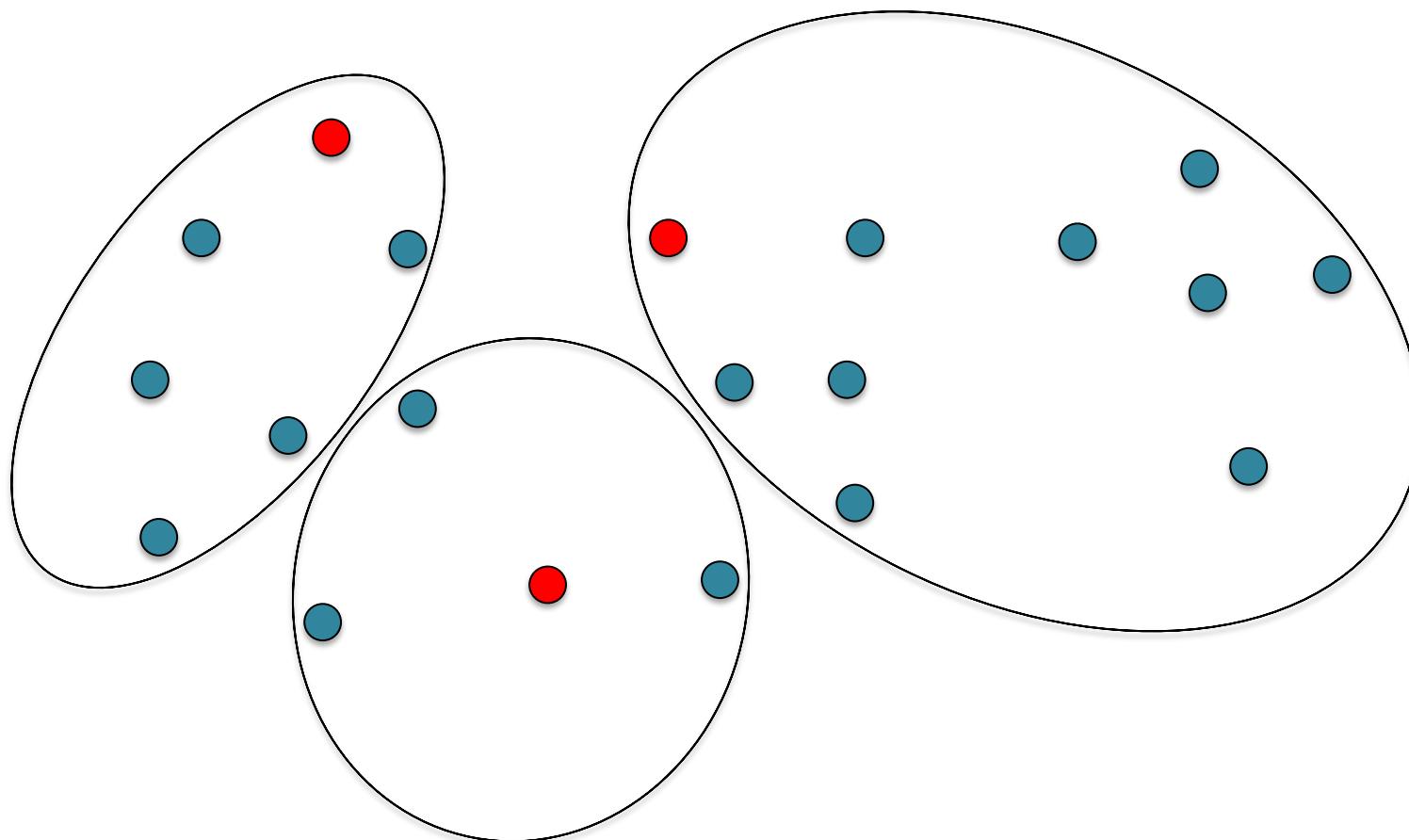
Clusterizarea K-means



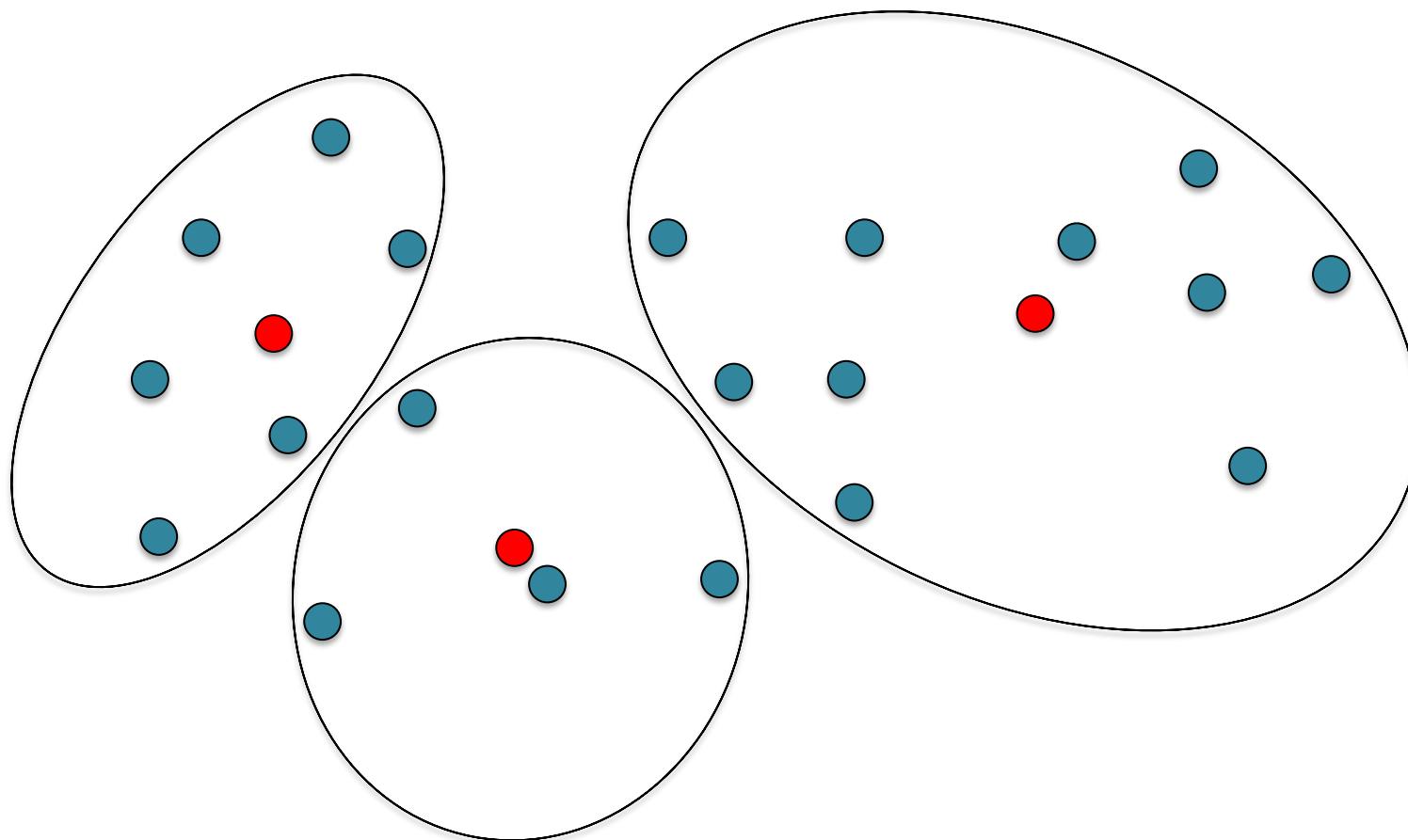
Clusterizarea K-means



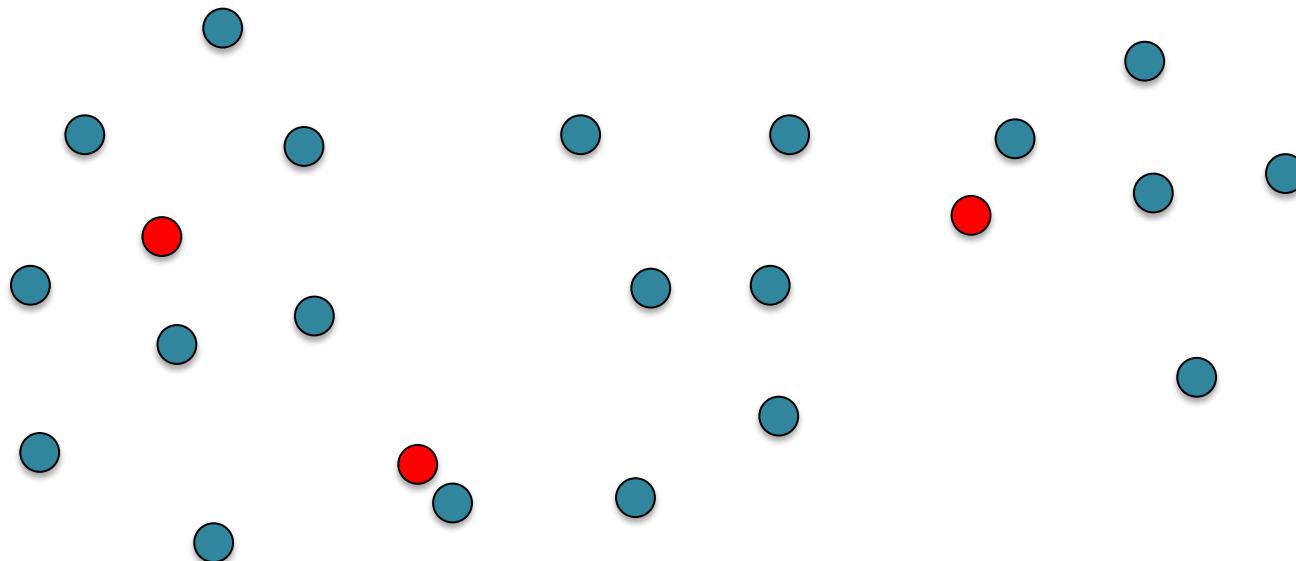
Clusterizarea K-means



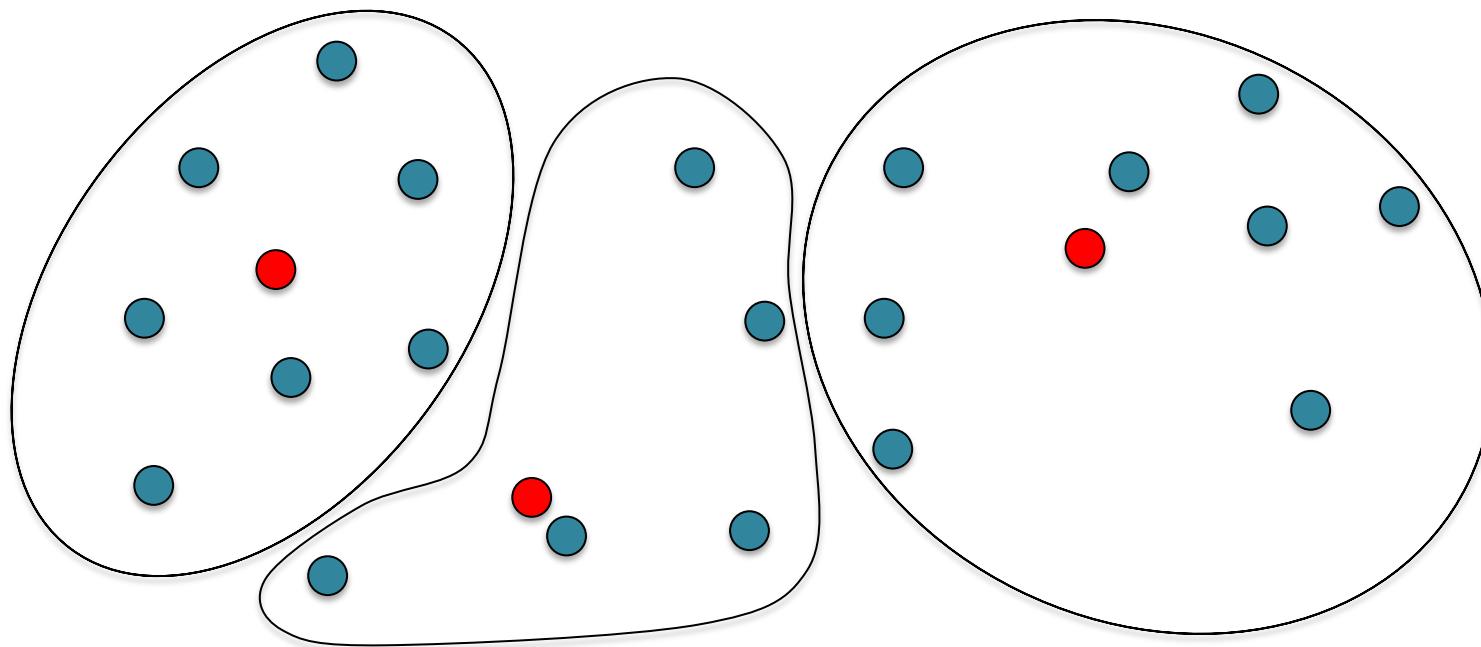
Clusterizarea K-means



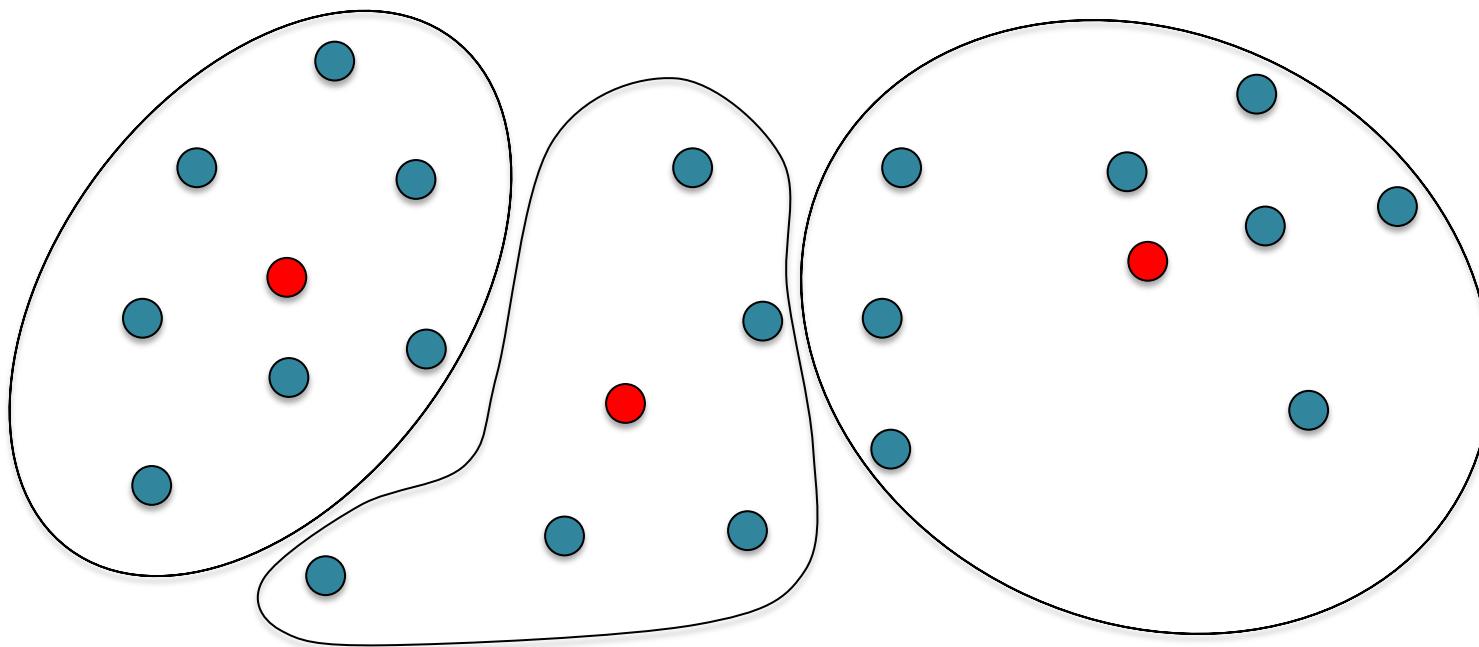
Clusterizarea K-means



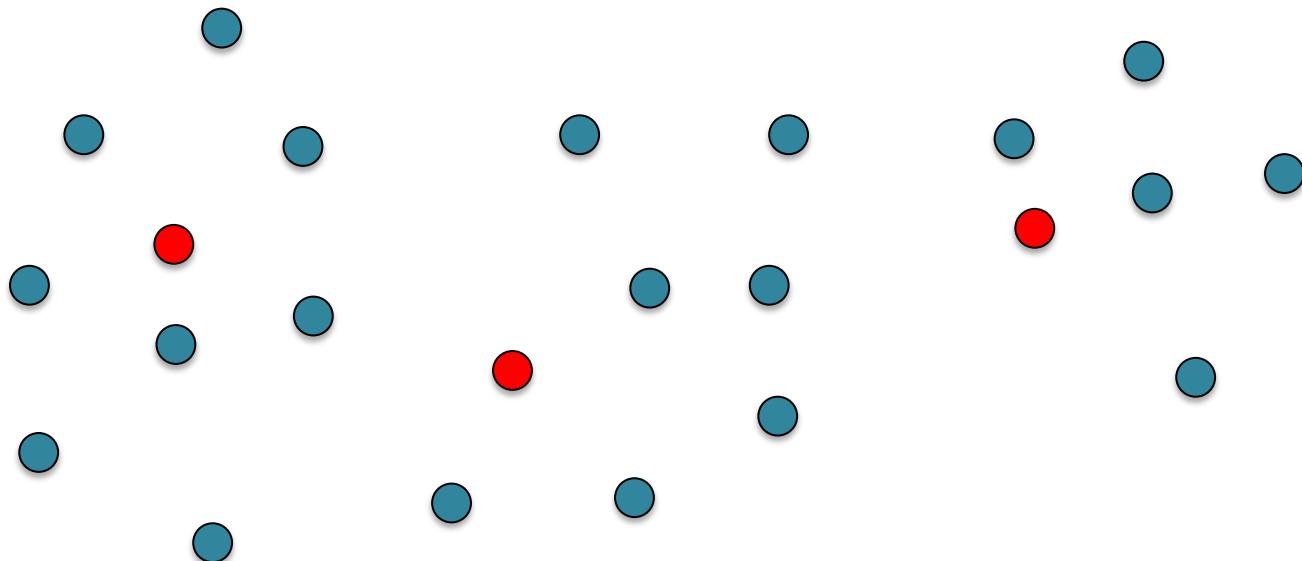
Clusterizarea K-means



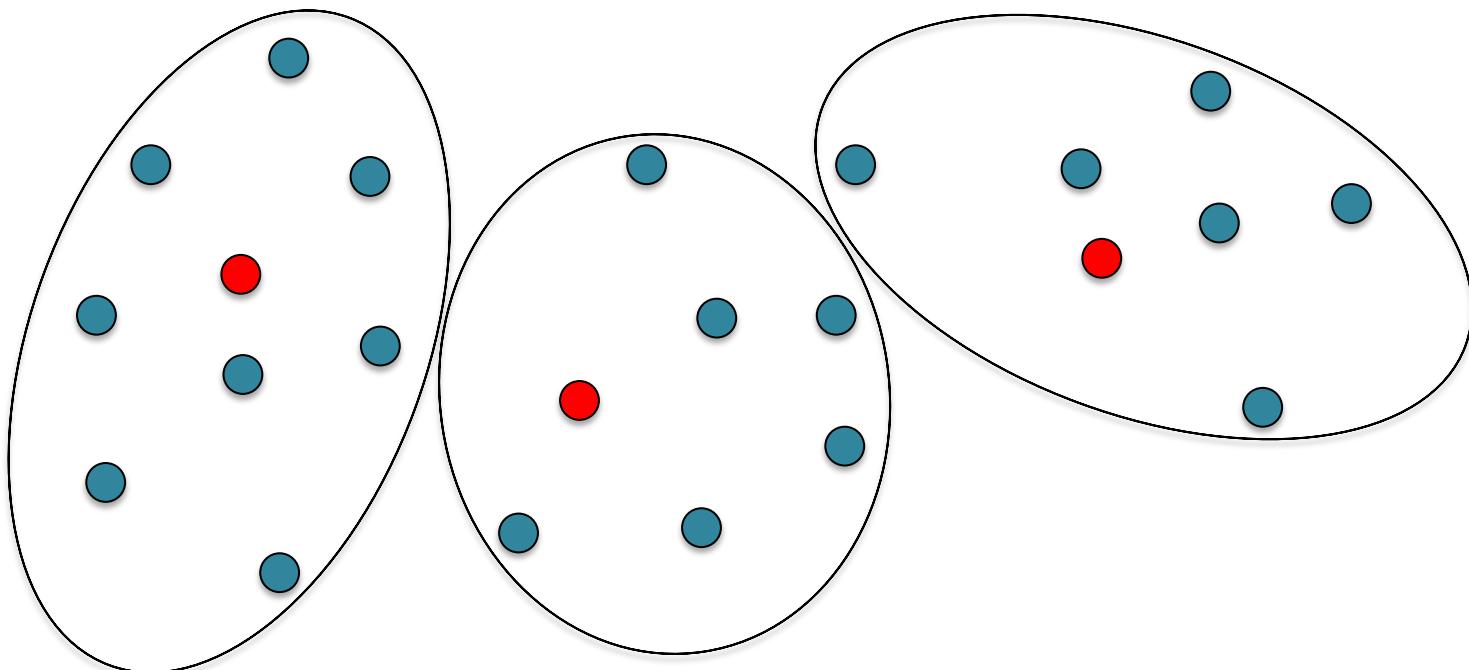
Clusterizarea K-means



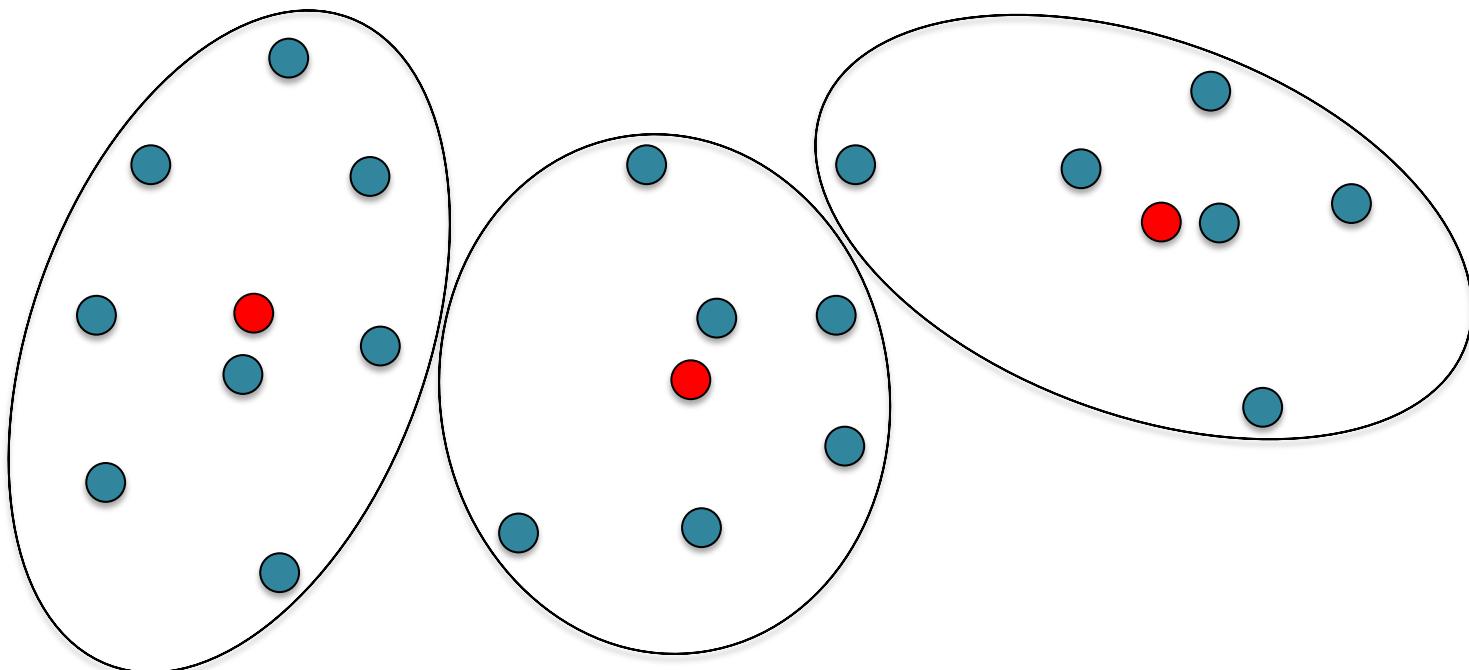
Clusterizarea K-means



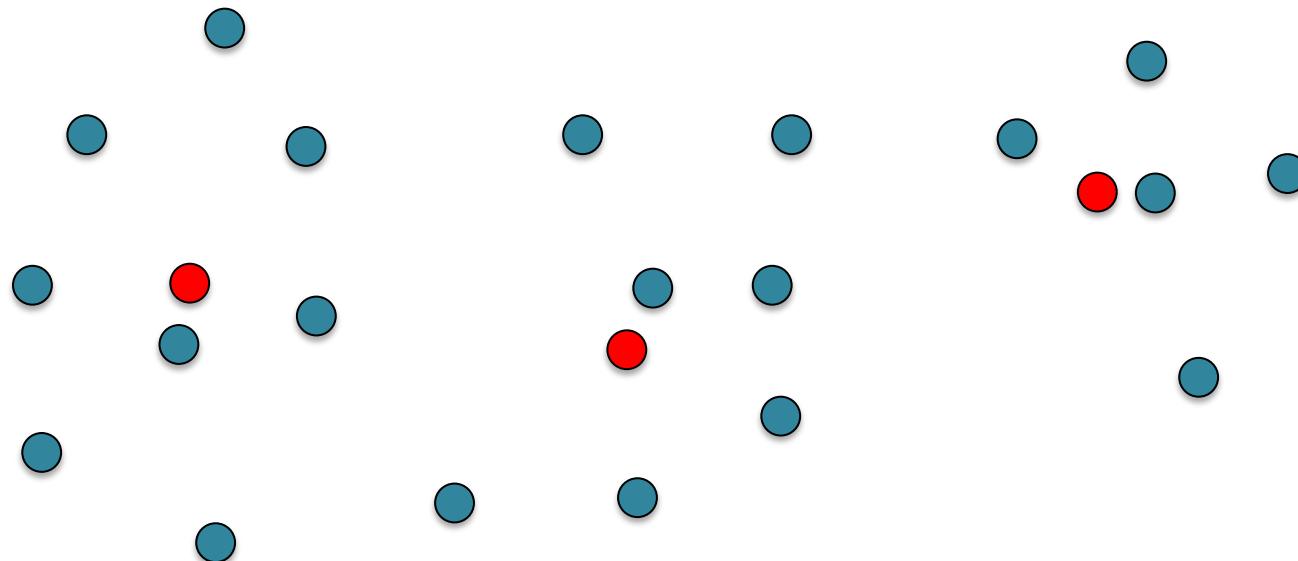
Clusterizarea K-means



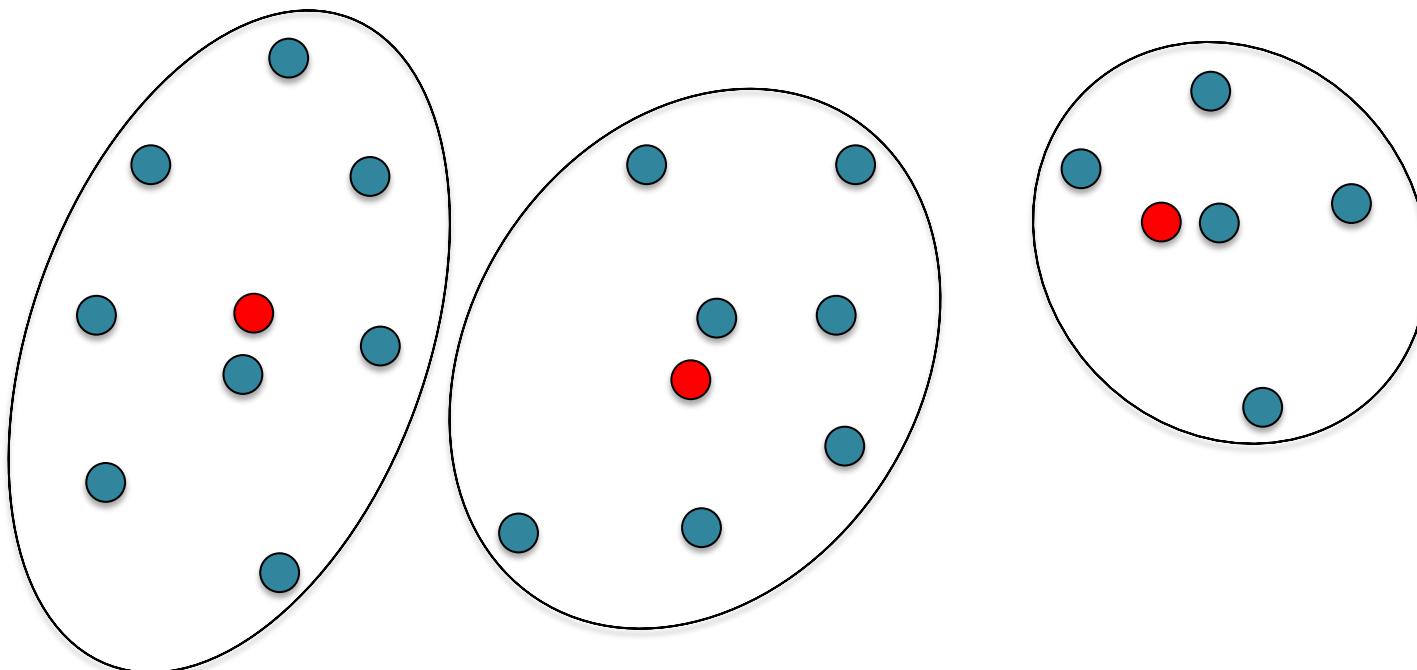
Clusterizarea K-means



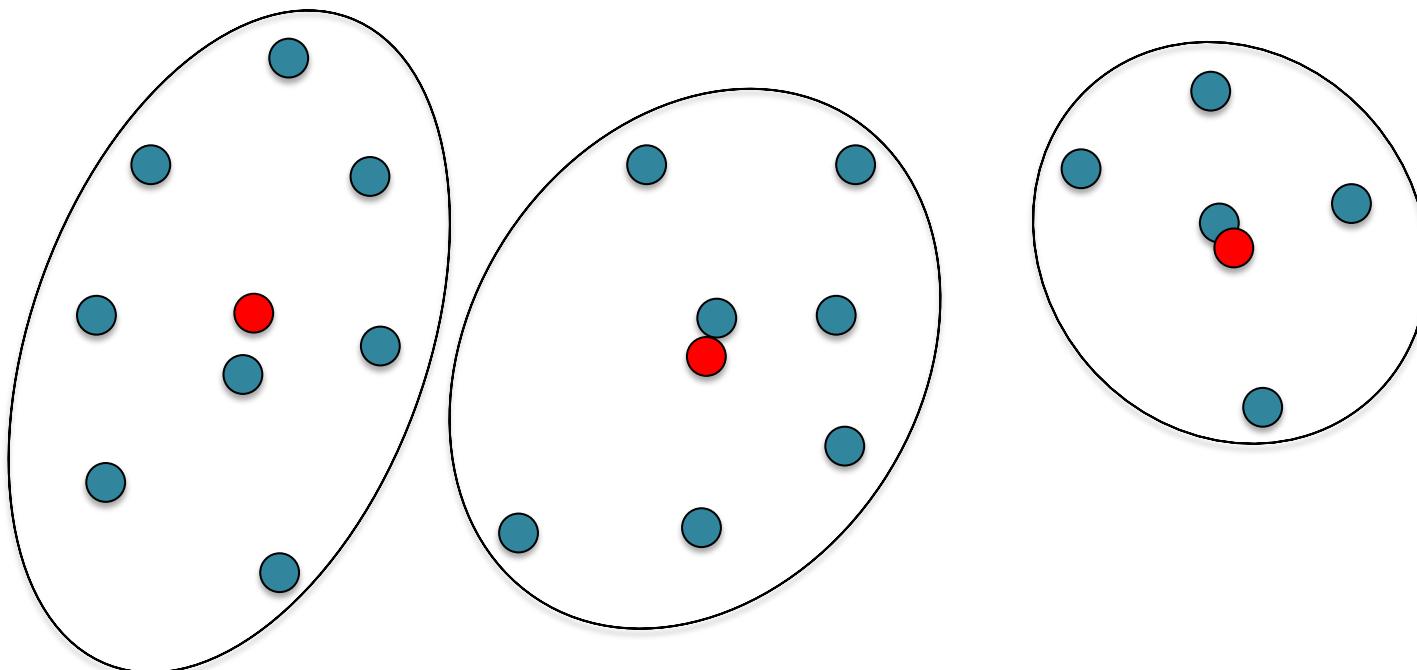
Clusterizarea K-means



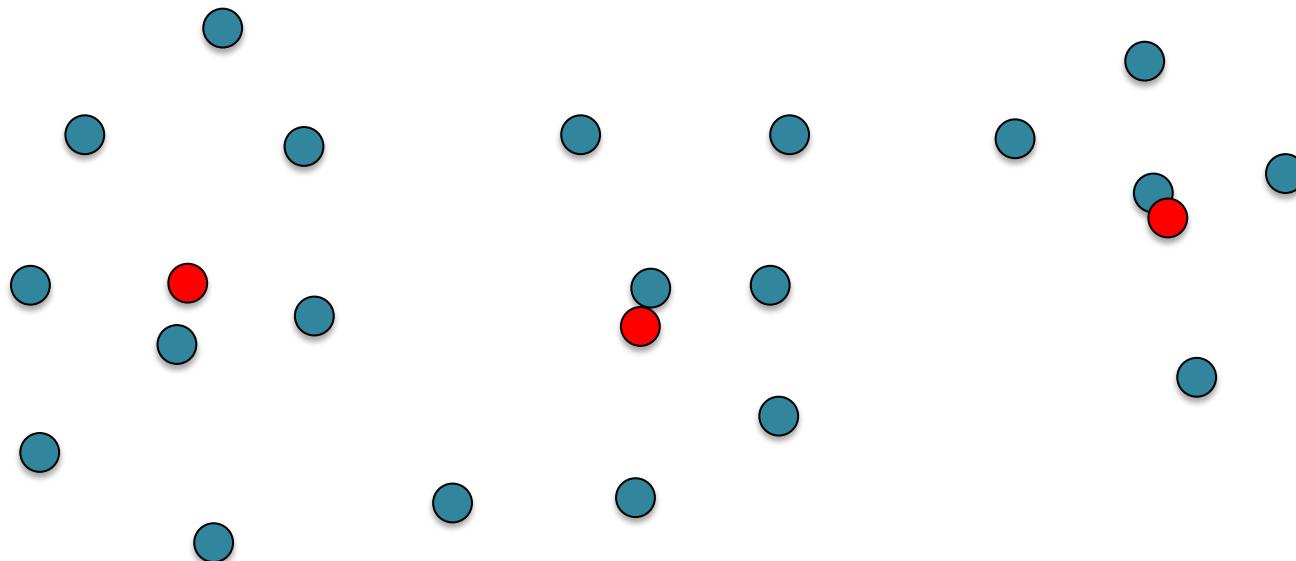
Clusterizarea K-means



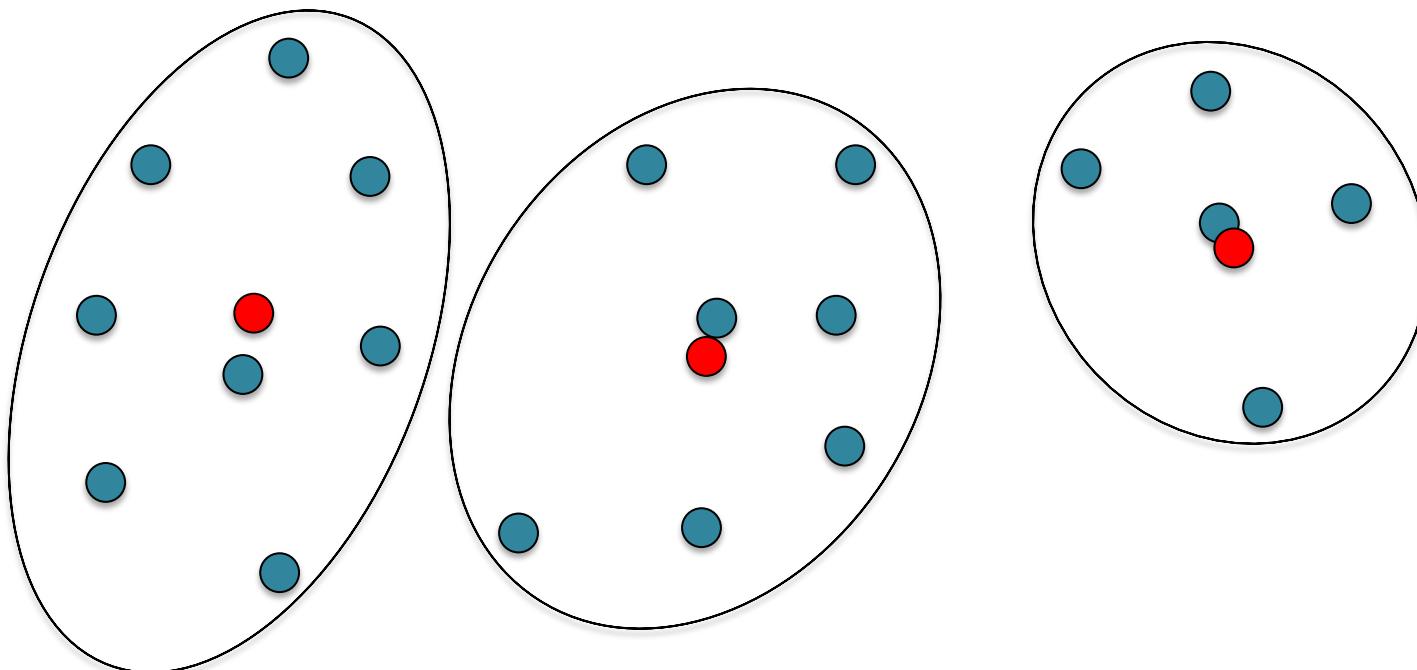
Clusterizarea K-means



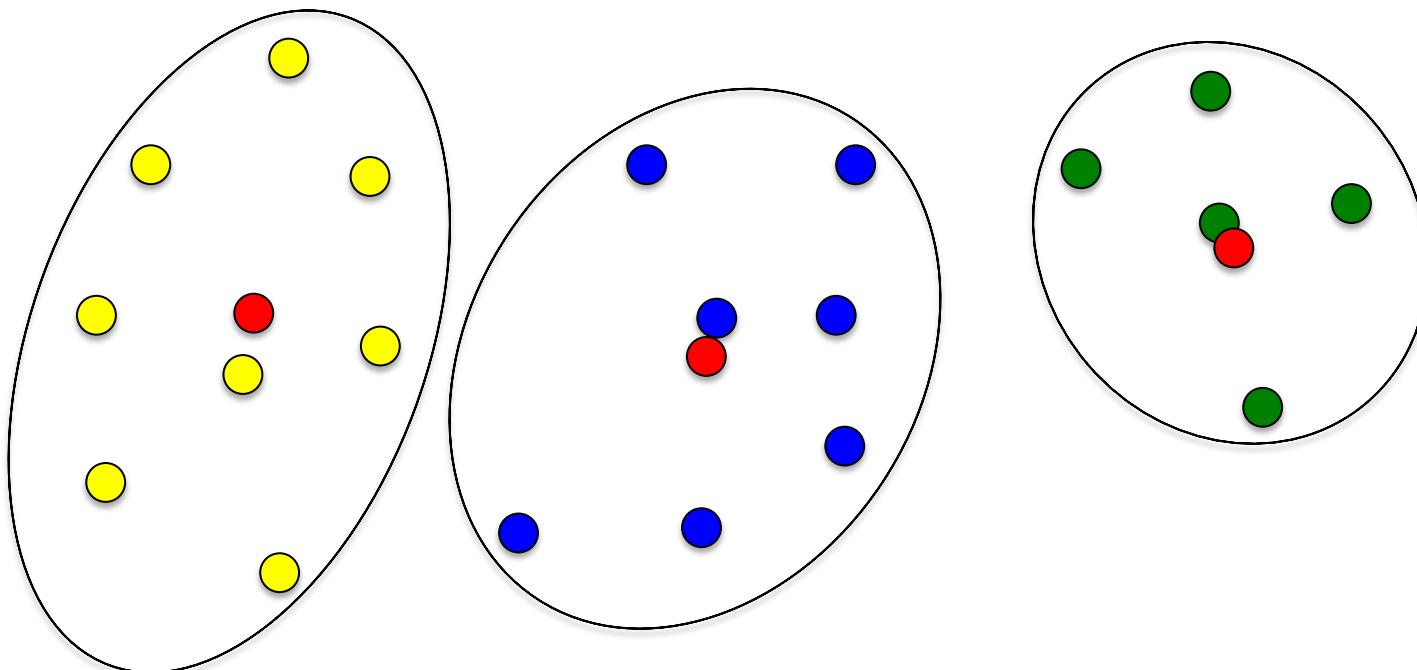
Clusterizarea K-means



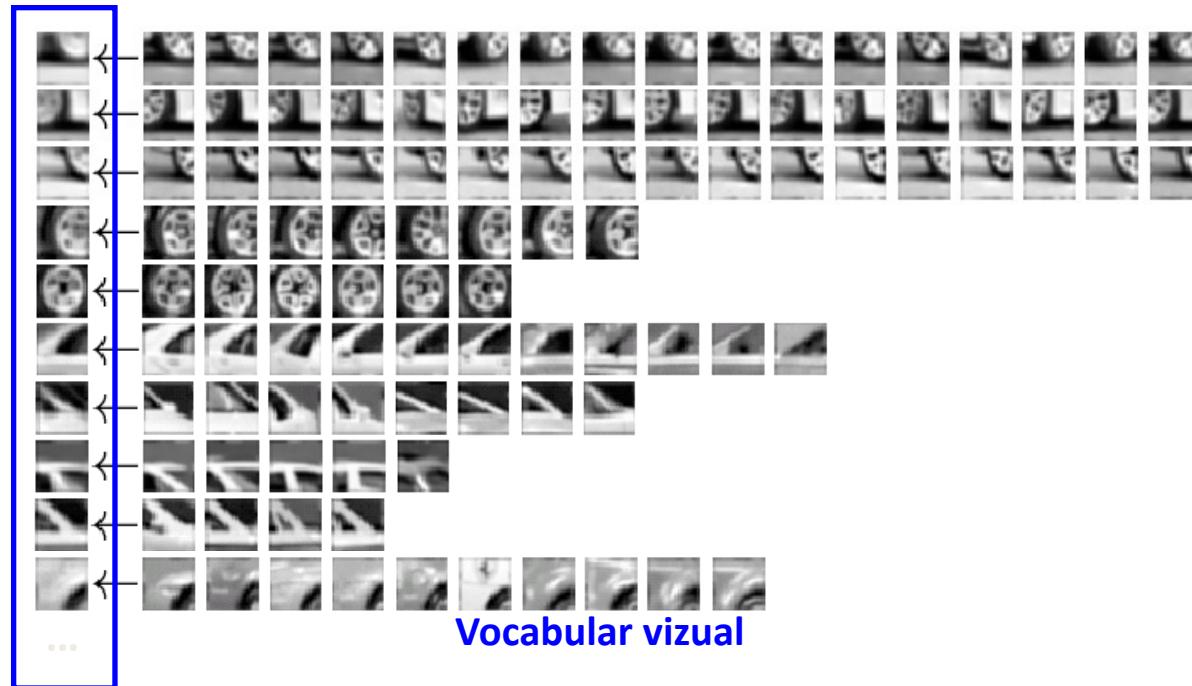
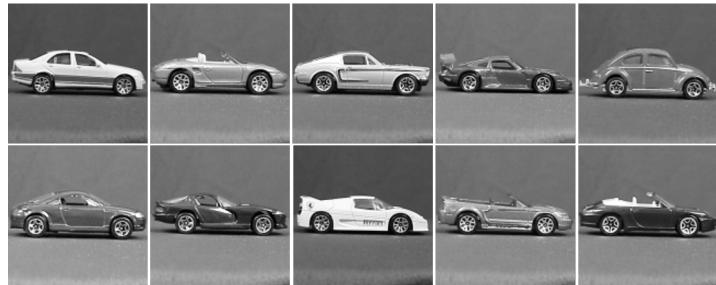
Clusterizarea K-means



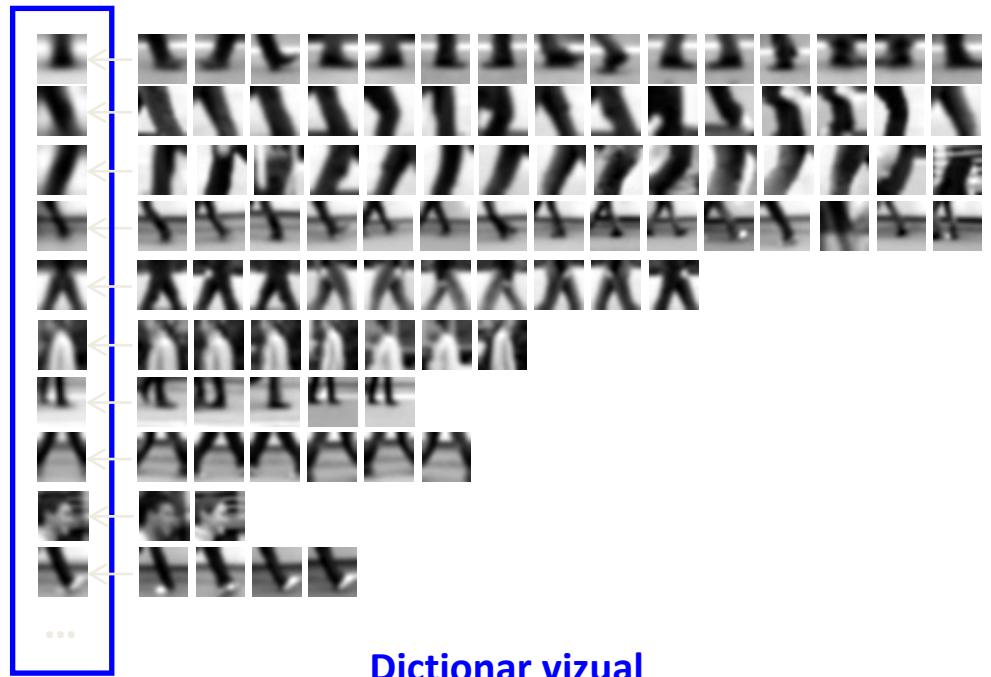
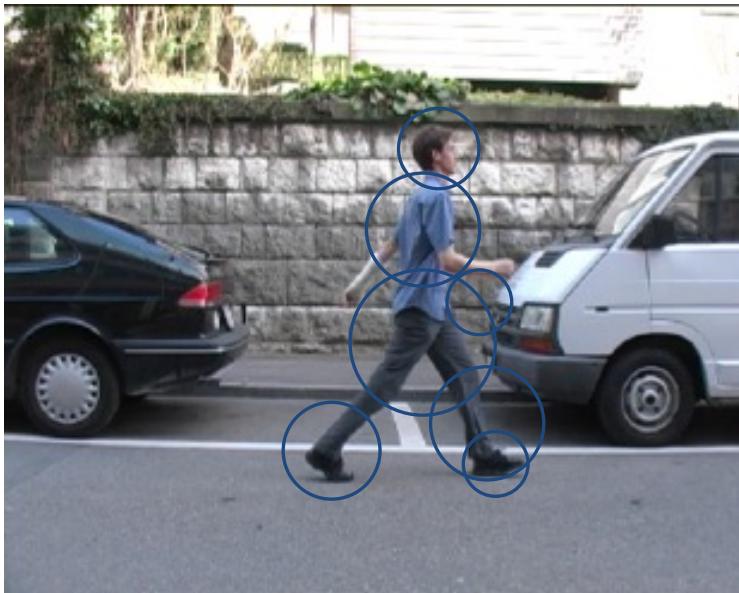
Clusterizarea K-means



Vocabular vizual - exemplu



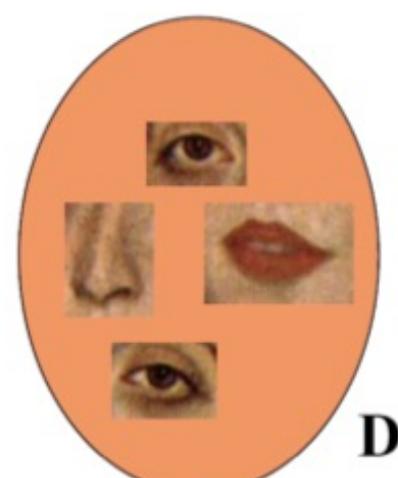
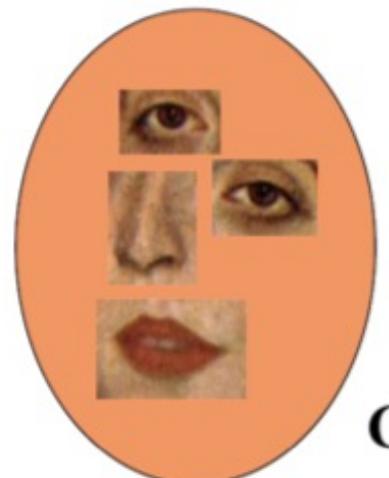
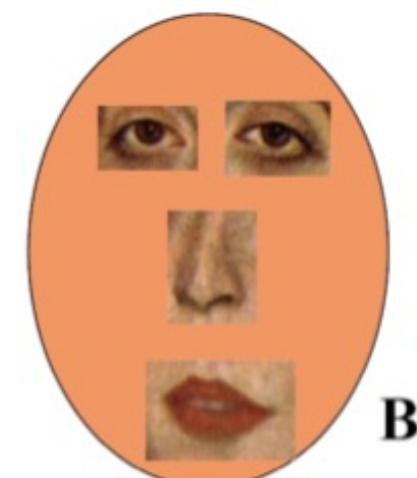
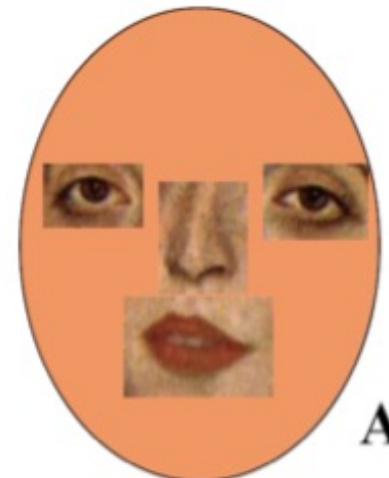
Vocabular vizual - exemplu



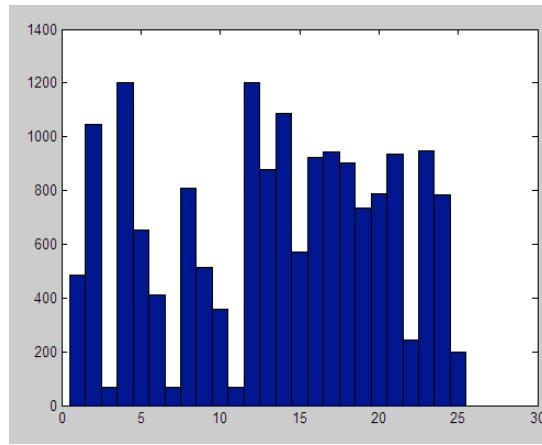
Dicționar vizual

Limitări ale modelului bag of visual words

- Reprezentare care nu ține cont de poziția în imagine a cuvintelor vizuale
- Avantaje?
- Dezavantaje?

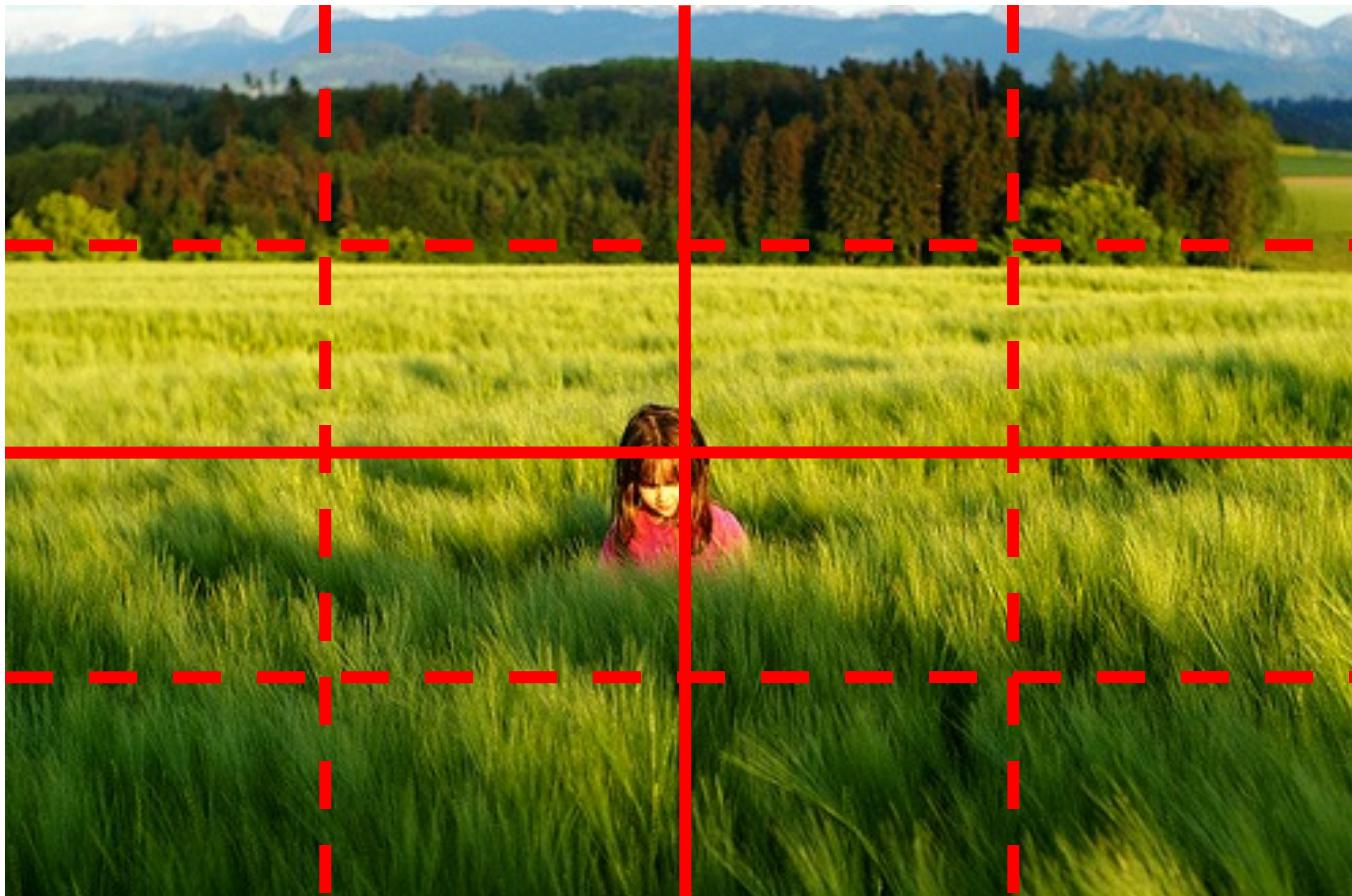


Poziția în imagine a caracteristicilor



Cele trei imagini au aceeași histogramă de culori.

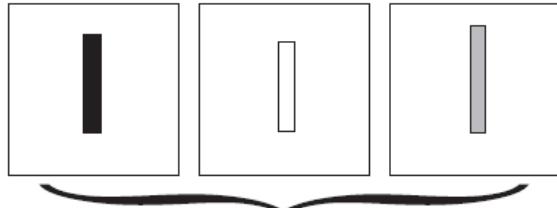
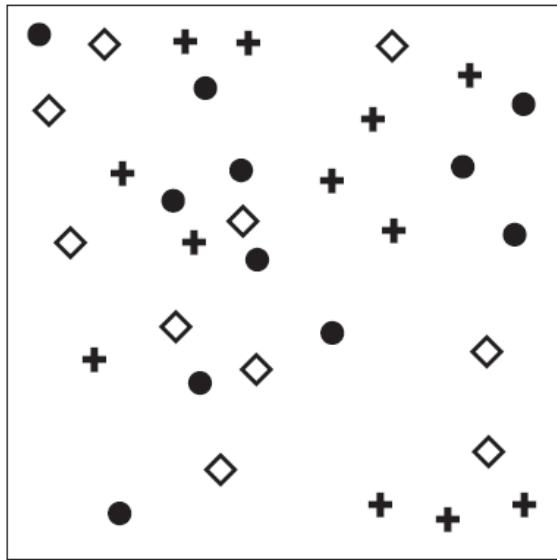
Piramidă spațială



Calculează histograma pentru fiecare sub-imagine

Piramidă spațială

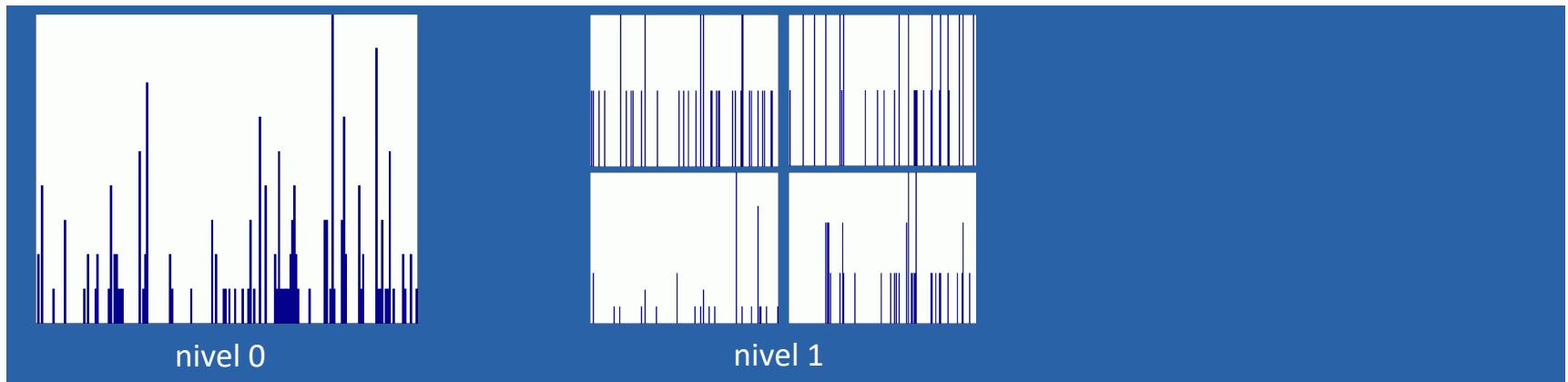
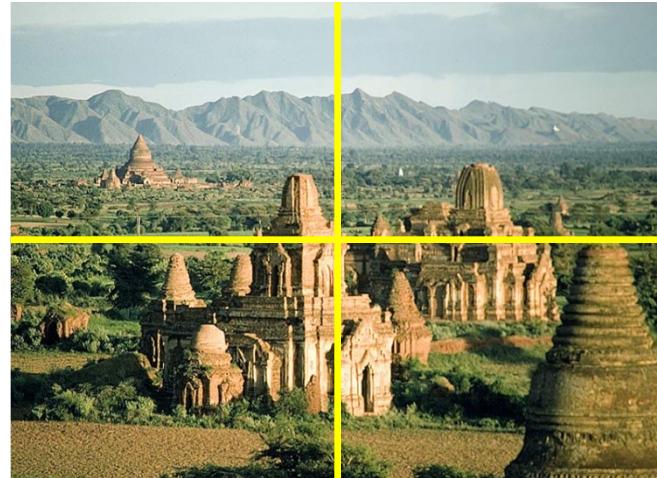
nivelul 0



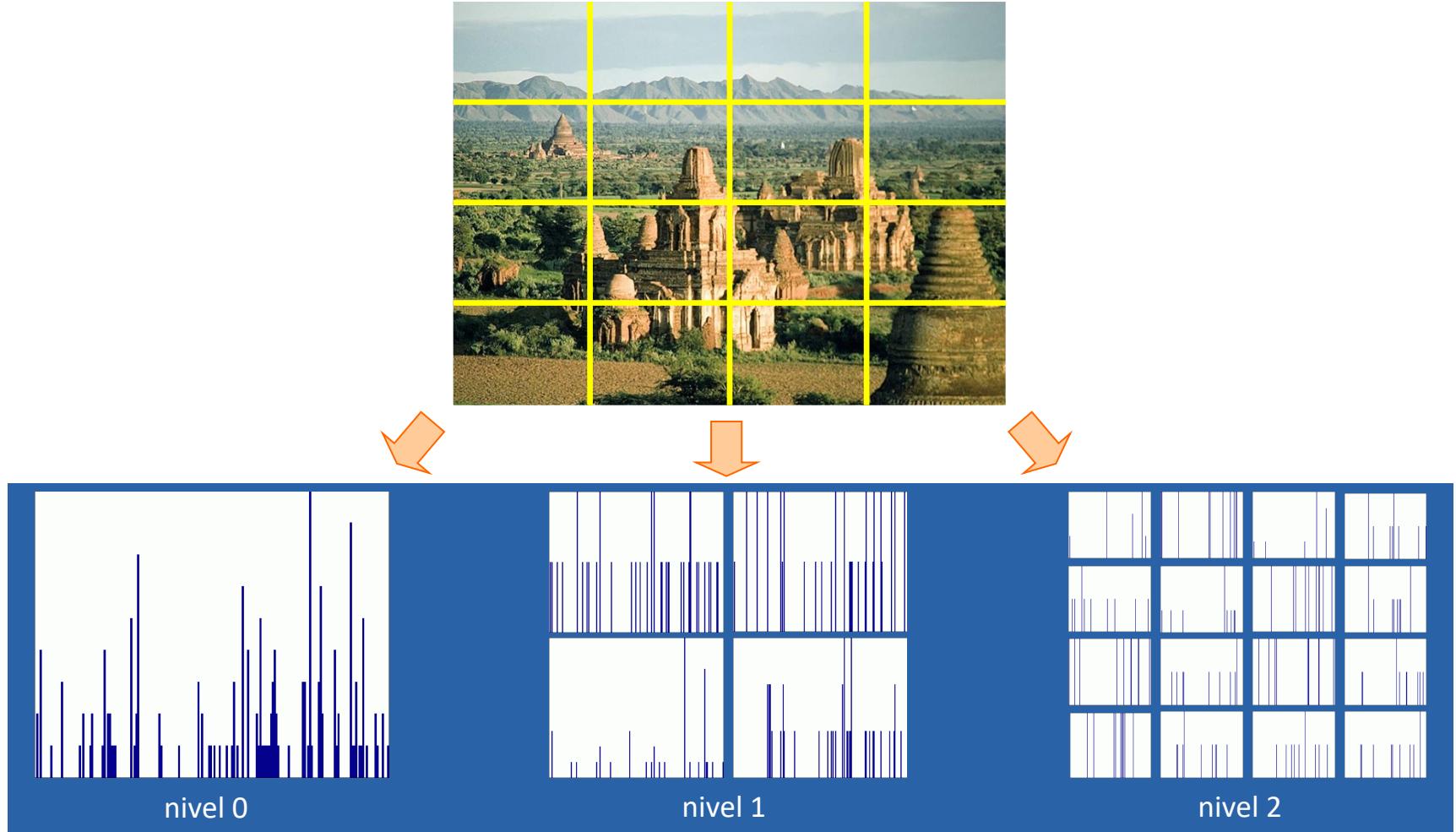
Piramidă spațială



Piramidă spațială



Piramidă spațială



Clasificarea imaginilor: modelul bag of visual words

Învățare

1. Extragă caracteristici (descriptori) pentru toate imaginile de învățare
2. Clusterizează descriptorii, centri clusterilor = visual words = “cuvinte vizuale”
3. Asignează fiecărui descriptor eticheta/index-ul unui “cuvânt vizual”
4. Reprezintă fiecare imagine printr-o histogramă normalizată de “cuvinte vizuale”
5. Învață un clasificator pe exemplele de învățare folosind histogramele normalizate drept caracteristici

Testare

1. Extragă descriptori din imaginea test, asignează fiecărui descriptor eticheta unui “cuvânt vizual”
2. Calculează histograma normalizată de “cuvinte vizuale”
3. Calculează eticheta corespunzătoare imaginii test sau un scor folosind clasificatorul învățat.

