



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale  
2007-2013



# Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

**AEACD**

## **10. Algoritmi de binarizare pentru documente moderne**

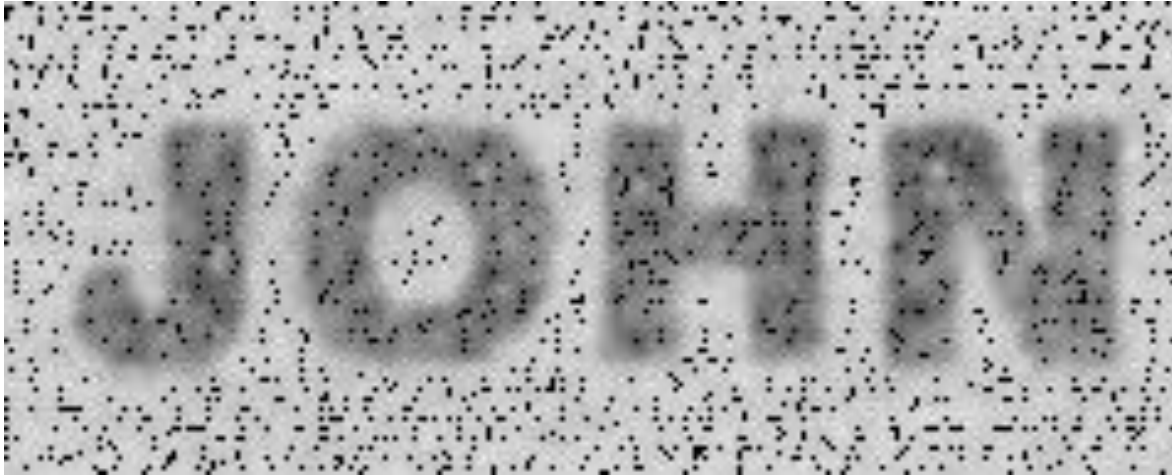
# Algoritmi pentru documente moderne

- Nu exista la fel de multe probleme
- Se pot aplica algoritmi generali cat si cei pentru documentele clasice
- Se poate folosi un algoritm specific pentru acest tip de documente ce rezolva si problema contrastului pentru cazul grayscale
- Algoritmul este alcatuit din mai multe parti

# Algoritmi pentru documente moderne

- Prima parte a algoritmului se ocupa cu descoperirea tuturor minimelor locale
- Astfel, de fiecare data cand se descopera un minim (un pixel cu o valoare grayscale mai mare decat ceilalti pixeli din vecinatate) acesta este marcat
- Se vor obtine o multime de puncte de minim local de la care se va porni reconstructia foregroundului

# Algoritmi pentru documente moderne



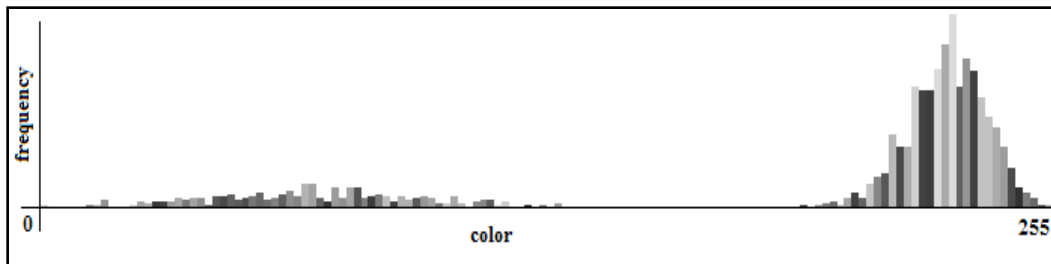
- In aceasta imagine puteti observa minimele locale descoperite

# Algoritmi pentru documente moderne

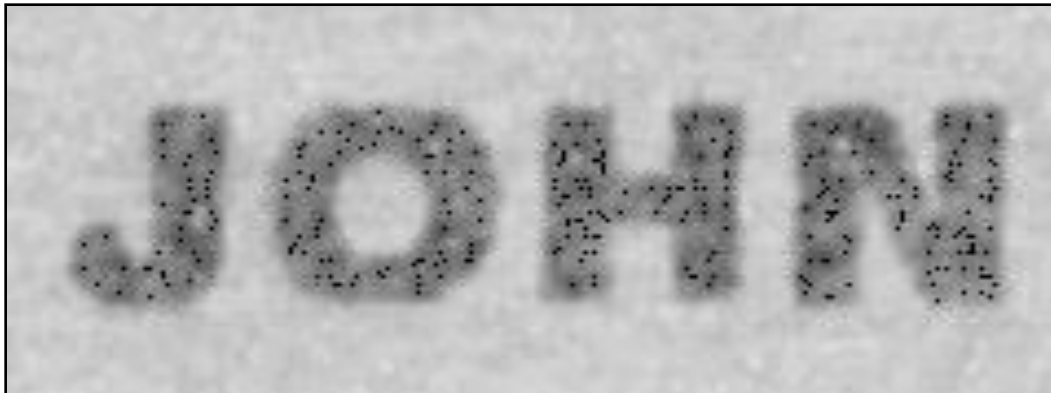
- In continuare aceste puncte trebuie filtrate pentru a obtine numai acei pixeli din foreground
- Pentru asta se realizeaza o histograma cu frecventa si valorile grayscale ale tuturor pixelilor
- Pe histograma se va observa cea mai importanta pauza a frecventelor valorilor pixelilor si astfel vom putea spune destul de sigur ca partea dreapta poate fi eliminata ca parte din background

# Algoritmi pentru documente moderne

- Histograma:



- Pixelii de minim local filtrati:



# Algoritmi pentru documente moderne

- Metoda de mai sus este expusa la anumite erori (ca de exemplu lipsa unei pauze pe histograma) si poate fi imbunatatita prin folosirea unui filtru triunghi intai si pe urma prin gasirea primului maxim local si primului minim local de la acesta
- Acestia vor reprezenta cel mai probabil backgroundul si tot ceea ce este cuprins intre aceste valori poate fi eliminat

# Algoritmi pentru documente moderne

- Pasul final se refera la reconstructia foregroundului pornind de la pixelii detectati
- Pentru fiecare minim local descoperit se parcurg toti pixelii in cele 4 directii: orizontal, vertical si oblic; de ambele parti ale pixelului; pana cand urmatorul pixel are o valoare grayscale mai mica decat cel curent
- Pentru fiecare pixel descoperit se asigneaza un scor ce reprezinta diferenta valorilor grayscale intre el si pixelul urmator

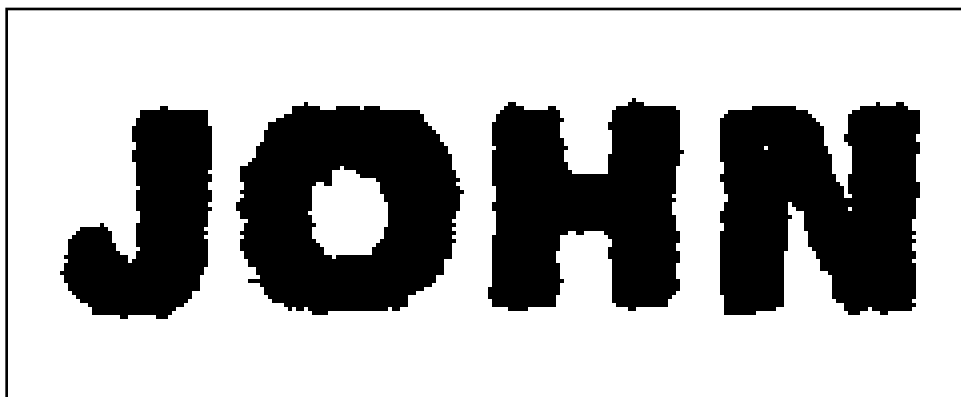


# Algoritmi pentru documente moderne

- Pe baza acestui scor se realizeaza cate o histograma pe fiecare din cele doua orientari ale celor 4 directii
- Pragul pentru reconstructie va fi selectat ca cea mai mare valoare din fiecare histograma
- Se unesc toti pixeli pana la primul scor cu acea valoare intalnit
- Acest pas se realizeaza pentru fiecare punct de minim local descoperit

# Algoritmi pentru documente moderne

- Rezultatul final:



- Acest algoritm poate fi imbunatatit prin alegerea unui alt scor de exemplu, insa pe ansamblu s-a comportat satisfactor, mai ales pe documente electronice
- De asemenea, facand mici modificari poate fi utilizat si pe imagini color, nu doar grayscale

# Algoritmi pentru documente generate electronic

- In general, nu exista mari probleme cu acest tip de documente
- Se pot aplica toti algoritmi deja enumerati, insa este de preferat sa se aleaga cel mai rapid pentru acest tip