

Introducere în PROCESAREA imaginilor

Radu-Daniel Vatavu

Procesarea imaginilor este un domeniu al inteligenței artificiale care se ocupă cu reprezentarea, reconstituirea, clasificarea, recunoașterea și analiza imaginilor cu ajutorul calculatorului.

Introducere

Sursa de proveniență a imaginilor poate fi un dispozitiv de achiziție (cameră video, scanner, captor radar), dar poate fi la fel de bine și o ecuație matematică, un ansamblu de date statistice etc.

Există două motive principale pentru care se apelează la procesarea imaginilor:

- îmbunătățirea calității unei imagini având drept scop o mai bună vizualizare pentru un operator uman. Aceasta poate însemna: reducerea zgomotului și a altor defecte care pot fi prezente în imagine (datorate, de exemplu, dispozitivului de achiziție), evidențierea unor zone de interes prin modificarea luminozității, a contrastului, accentuarea muchiilor etc.;
- extragerea de informații dintr-o imagine, informații care pot reprezenta intrarea pentru un sistem automat de recunoaștere și clasificare. Aceste informații pot fi: diferite distanțe și relații dintre obiectele prezente în imagine, momente statistice, parametri geometrici (arie, perimetru, circularitate), coeficienți *Fourier* etc.

Exemple clasice de aplicații pentru procesarea imaginilor (și recunoașterea formelor) includ: recunoașterea caracterelor, recunoașterea amprentelor, prelucrarea imaginilor medicale, a imaginilor satelit.

Noțiuni de bază

În cele ce urmează se va considera drept reprezentare a unei imagini un tablou bidimensional de numere întregi (*pixeli*). Valoarea fiecărui pixel descrie nivelul de strălucire sau culoarea acestuia. În cel mai simplu caz, al imaginilor binare, pentru reprezentarea fiecărui pixel este folosit un singur bit. În cazul imaginilor cu niveluri de gri, valoarea fiecărui pixel reprezintă strălucirea acestuia.

Cel mai comun format pentru aceste imagini are la bază reprezentarea pixelilor pe 8 biți. Astfel, gama de valori posibile este 0..255, 0 codificând culoarea **negru**, 255 **alb**, iar valorile intermediare reprezentând nuanțe (niveluri) de gri. În cazul imaginilor color, pot fi folosite diferite sisteme de culori (*RGB*, *HSI*, *CMY* etc.).

Modele de culori

Modelul *RGB* este un model aditiv, având culorile fundamentale roșu, verde și albastru, celelalte culori constituindu-se din anumite procente ale culorilor de bază. Culoarea alb are prezente toate culorile fundamentale, iar culoarea negru se caracterizează prin absența lor. Modelul *RGB* este adesea reprezentat sub forma unui cub unitate, ca în figura 1:

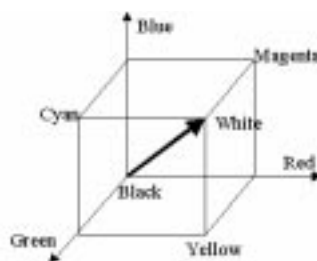


Figura 1

Originea (0, 0, 0) este considerată culoarea **negru**, iar colțul opus (1, 1, 1) corespunde culorii **alb**. Diagonala care unește aceste două puncte conține nivelurile de gri obținute prin utilizarea aceluiași procent din culorile de bază.

O altă metodă de clasificare a culorilor este prin **nuanțare** (*Hue*), **saturație** (*Saturation*) și **intensitate** (*Intensity*), rezultând modelul *HSI*.



Nuanța reprezintă poziția relativă a culorii în spectru și, în modelul *HSI*, corespunde unghiului pe care îl face culoarea în cercul de culori. Domeniul de nuanțare se întinde de la 0 la 360 de grade.

Saturația reprezintă puritatea culorii. Valoarea saturației se exprimă în procente și variază de la 0 (fără culoare) până la 100 (culoare pură, așa cum este definită de valoarea nuanțării).

Ultimul parametru reprezintă intensitatea culorii. O reprezentare simbolică a modelului *HSI* este dată în figura 2. Nuanța determină culoarea (unghi = 60° - *Yellow*, 180° - *Cyan* etc.) Saturația este măsurată în procente de la centrul cercului de bază la suprafața conurilor.

Intensitatea se măsoară pe linia *Black-White*.

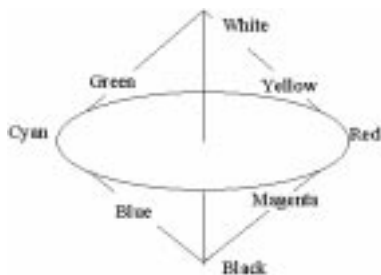


Figura 2

Modelul *CMY* are la bază culorile *Cyan*, *Magenta* și *Yellow*, considerate culori secundare față de culorile *Red*, *Green*, *Blue*.

Modelul este substractiv, culoarea *alb* fiind dată de absența culorilor *CMY*, iar culoarea *negru* fiind caracterizată de prezența acestora.

În continuare sunt date câteva modalități de conversie între tipurile de modele prezentate:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix} \quad \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} \frac{R}{2} - \arctan\left(\frac{2R-G-B}{\sqrt{3}(G-B)}\right) + \pi \\ 1 - \frac{3}{(R+G+B)} \arctan\left(\frac{R-G-B}{\sqrt{3}}\right) \\ \frac{R+G+B}{3} \end{bmatrix}$$

Vecinătăți și conectivitate

Începând cu această secțiune ne vom referi prin f_{ij} la valoarea pixelului aflat la coordonatele (i, j) din matricea care reprezintă imaginea (unde i reprezintă linia, iar j coloana). Vom considera că matricea este de dimensiune $M \times N$. De asemenea, dacă nu se precizează altfel, discuțiile vor avea în vedere o imagine cu niveluri de gri (considerăm reprezentarea pe 8 biți, deci cu 256 de valori posibile). Aceasta nu reprezintă o limitare, întrucât o imagine color poate fi descompusă după culorile fundamentale (R, G, B) în trei imagini separate, fiecare astfel de imagine putând fi tratată ca alcătuită din niveluri de gri.

Un pixel p de coordonate (i, j) prezintă patru vecini în punctele: $(i-1, j)$, $(i+1, j)$, $(i, j-1)$, $(i, j+1)$. Vom nota acest ansamblu (vecinătate) de pixeli cu $N_4(p)$. Pixelii diagonali

de coordonate: $(i-1, j-1)$, $(i-1, j+1)$, $(i+1, j+1)$, $(i+1, j-1)$ alcătuiesc un nou ansamblu notat $N_D(p)$. În final, ansamblurile $N_4(p)$ și $N_D(p)$ reunite, formează ansamblul notat $N_8(p)$.

Având în vedere ansamblurile (vecinătățile) definite mai sus, se pot considera două tipuri de conectivitate:

- **4-conectivitate** - doi pixeli, p_1 și p_2 , de coordonate (i_1, j_1) și (i_2, j_2) sunt în relație de 4-conectivitate dacă (i_1, j_1) se află în ansamblul $N_4(p_2)$ sau echivalent, dacă (i_2, j_2) se află în ansamblul $N_4(p_1)$;
- **8-conectivitate** - doi pixeli, p_1 și p_2 , de coordonate (i_1, j_1) și (i_2, j_2) sunt în relație de 8-conectivitate dacă (i_1, j_1) se află în ansamblul $N_8(p_2)$ sau echivalent, dacă (i_2, j_2) se află în ansamblul $N_8(p_1)$.

Se numește drum de lungime n mulțimea de pixeli $D_n = \{p_1, p_2, \dots, p_n\}$ pentru care fiecare pixel p_i se află în relație de conectivitate cu pixelul următor p_{i+1} , $i=1 \dots n-1$.

Doi pixeli p și q sunt **conectați** dacă există un drum D_n de la p la q .

Un set de pixeli formează o **regiune** dacă orice doi pixeli aparținând setului sunt conectați.

Metrici

Distanțele dintre pixeli poartă denumirea de **metrice**. Fie P mulțimea pixelilor unei imagini. O **metrică**, în cazul unei imagini, este o funcție $d: P \times P \rightarrow R$ care îndeplinește următoarele condiții:

$$\begin{aligned} d(p_1, p_2) &\geq 0 \quad \forall p_1, p_2 \in P. \quad d(p_1, p_2) = 0 \Leftrightarrow p_1 = p_2 \\ d(p_1, p_2) &= d(p_2, p_1) \quad \forall p_1, p_2 \in P \\ d(p_1, p_3) &\leq d(p_1, p_2) + d(p_2, p_3) \quad \forall p_1, p_2, p_3 \in P \end{aligned}$$

Câteva exemple de metrice utilizate frecvent sunt:

- **distanța euclidiană:**

$$d_e(p_1, p_2) = \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2},$$

unde p_1 și p_2 sunt pixeli de coordonate (i_1, j_1) , respectiv (i_2, j_2) .

- **distanța Manhattan (city block metric):**

$$d_m(p_1, p_2) = |i_1 - i_2| + |j_1 - j_2|.$$

În baza acestei distanțe putem da o definiție riguroasă pentru ansamblul $N_4(p)$, astfel:

$$N_4(p) = \{q \in P / d_m(p, q) \leq 1\},$$

- **distanța jocului de șah (chessboard metric):**

$$d_c(p_1, p_2) = \max\{|i_1 - i_2|, |j_1 - j_2|\}$$

Având această funcție, se poate da o nouă definiție pentru ansamblul $N_8(p)$, astfel:

$$N_8(p) = \{q \in P / d_c(p, q) \leq 1\}.$$

Histograma nivelurilor de gri

Histograma nivelurilor de gri a unei imagini reprezintă o funcție care asociază fiecărui nivel de gri prezent în imagine frecvența sa de apariție.

Dacă luăm în considerare o imagine f care prezintă un număr maxim G de niveluri de gri, atunci funcția histogramă poate fi definită astfel:

$$h: \{0, 1, \dots, G-1\} \rightarrow \{0, 1, \dots, N \times M\}$$

$$h(k) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f_{i,j} = k], \text{ unde } [x = y] = \begin{cases} 1, & \text{pentru } x = y \\ 0, & \text{pentru } x \neq y \end{cases}$$

Un exemplu de histogramă este dat în figura următoare:

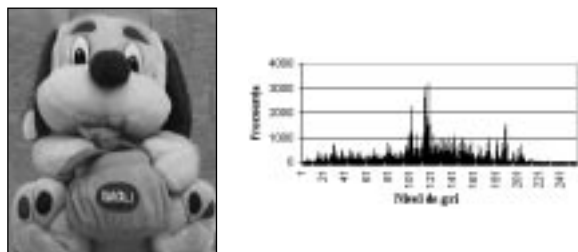


Figura 3: Imagine cu 256 de niveluri de gri și histograma nivelurilor de gri

Histograma nu oferă nici o informație cu privire la locația pixelilor sau la poziția relativă a acestora, însă, dispunând de histograma unei imagini, nivelurile de gri pot fi manipulate pentru a îmbunătăți contrastul, luminozitatea, etc. De exemplu, o deplasare a nivelurilor de gri din histogramă spre stânga/dreapta conduce la o diminuare/creștere a luminozității.

Se poate observa din figura 3 o concentrare a valorii unor pixeli în jurul unor anumite valori (de exemplu 120), reprezentând regiuni cu aceeași luminozitate sau luminozități foarte apropiate (cum ar fi fondul imaginii). Acest aspect este foarte util în procesul de segmentare permițând o separare rapidă a obiectelor de fundal în funcție de luminozitatea acestora.

În cazul imaginilor color, folosind modelul RGB , se pot realiza histogramme pentru fiecare culoare de bază în parte cu păstrarea validității celor discutate mai sus.

Preprocesarea imaginilor

Preprocesarea are drept scop realizarea unei îmbunătățiri a imaginii prin punerea în evidență a anumitor regiuni, schimbarea luminozității, detecția muchiilor etc., pregătind imaginea în vederea operației de segmentare.

Transformările aplicate în această etapă unei imagini pot fi de două tipuri:

- transformări punctuale (modificarea valorii unui pixel se face independent de vecinătatea acestuia);
- transformări locale (noua valoare a pixelului depinde de valorile pixelilor înconjurători).

Transformări punctuale

Din cadrul transformărilor punctuale vom avea în vedere manipularea contrastului prin transformarea nivelurilor de gri. În acest sens, o transformare punctuală poate fi privită ca o funcție de transfer T definită pe mulțimea nivelu-

rilor de gri originale, care atribuie fiecărui pixel o nouă valoare în funcție de luminozitatea sa. Exemple de funcții de transfer sunt date în următoarea figură:

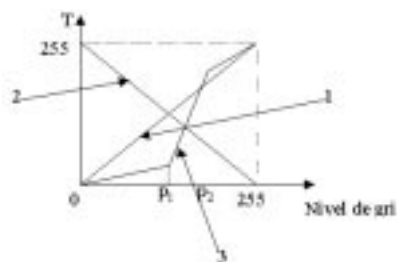


Figura 4

- funcția identitate: $T(p) = p, p = 0 \dots 255$;
- funcția inversă: $T(p) = 255 - p, p = 0 \dots 255$;
- o funcție de tipul:

$$T(p) = \begin{cases} a_1 p + b_1, & \text{pentru } p < p_1 \\ a_2 p + b_2, & \text{pentru } p_1 \leq p \leq p_2 \\ a_3 p + b_3, & \text{pentru } p > p_2 \end{cases}$$

Funcția 3 prezentată poate fi folosită pentru accentuarea contrastului dintre nivelurile de gri cuprinse între p_1 și p_2 , dacă panta a_2 este superioară valorilor a_1 și a_3 .

Egalizarea histogramei

Transformările nivelurilor de gri au drept scop o mai bună vizualizare a imaginii de către un operator uman. O tehnică din cadrul acestor transformări, care dă rezultate bune în practică, poartă denumirea de **egalizarea histogramei** și constă în distribuirea nivelurilor de gri existente pe întreaga gamă posibilă de valori.

Această tehnică este foarte utilă în cazul unor imagini ale căror niveluri de gri sunt localizate într-un interval restrâns, iar "împrăștierea" lor pe întreaga scală conduce la o vizualizare sensibil mai bună. Funcția de transfer utilizată în acest caz are la bază histograma nivelurilor de gri, fiind reprezentată de curba cumulată a frecvențelor nivelurilor de gri.

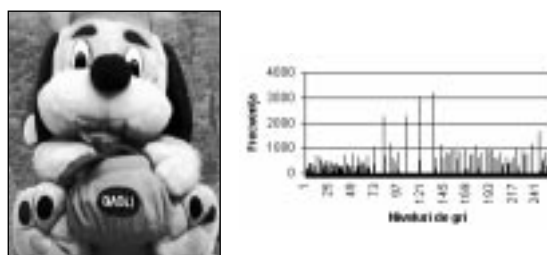


Figura 5: Imaginea după egalizarea histogramei și histograma după egalizare

În figura 5 este prezentată imaginea obținută prin aplicarea metodei **egalizării histogramei** originalului din figura 3.

Se observă accentuarea contrastului ce permite o mai clară observare a detaliilor, punerea în evidență a variațiilor





lor minore dintre regiuni care trec neobservate în imaginea inițială.

Concret, metoda egalizării histogramei constă în înlocuirea fiecărui nivel de gri k din imaginea originală cu valoarea frecvențelor cumulate până la acest nivel (valoare scalată, bineînțeles, în intervalul 0 ... 255).

Trebuie notat faptul că egalizarea histogramei poate fi realizată doar pentru o anumită regiune de interes și nu pentru toată imaginea (de exemplu, o zonă determinată de un dreptunghi).

Transformări locale

Aceste transformări țin cont de o anumită vecinătate a pixelului al cărui nivel de gri va fi modificat. De obicei, vecinătățile cu care se lucrează sunt pătratice, de dimensiuni 3×3 , 5×5 , 7×7 , în general de forma $(2p+1) \times (2p+1)$, centrate pe pixelul supus transformării.

Transformările locale, în funcție de scopul lor, se pot clasifica în două categorii:

- transformări care vizează *netezirea imaginii* (*image smoothing*) sau a unor regiuni din aceasta, având drept scop reducerea zgomotului sau a altor defecte ale imaginii;
- transformări care vizează *detectia muchiilor* (*edge detection*), bazate pe derivatele funcției asociată imaginii.

O altă clasificare a transformărilor locale are în vedere proprietățile de liniaritate ale acestora. Astfel, putem vorbi despre **transformări liniare** și **transformări neliniare**.

Transformările liniare se caracterizează prin faptul că valoarea rezultată reprezintă o combinație liniară a nivelurilor de gri din vecinătatea utilizată pentru fiecare pixel. Contribuția fiecărui pixel din vecinătate în rezultatul final este ponderată cu o anumită valoare, astfel:

unde (i^0, j^0) sunt coordonatele pixelului asupra căruia se realizează transformarea, f reprezintă imaginea inițială iar, f^{out} imaginea finală.

Ecuția de mai sus poate fi scrisă ca un **produs de convoluție** dintre imaginea originală f și nucleul (matricea) w , astfel:

Reducerea zgomotului unei imagini

Metodele privind reducerea zgomotului prezent în imagine (*filtrare*) pot fi liniare sau neliniare, în acord cu discuțiile de mai sus.

Filtrarea liniară se bazează pe înlocuirea nivelului de gri al unui pixel cu media (ponderată sau nu) a valorilor pixelilor din vecinătatea sa.

O metodă simplă care poate rezolva acest deziderat este utilizarea unor nuclee de convoluție de tipul:

Se observă că prin aplicarea nucleului w_1 imaginii originale, valoarea fiecărui pixel este înlocuită cu media valorilor pixelilor dintr-o vecinătate 3×3 centrată pe pixelul respectiv. Desigur, pot fi folosite vecinătăți de dimensiuni mai mari: 5×5 , 7×7 etc. În cazul celorlalte două exemple se atribuie o pondere mai mare pixelului central în calculul mediei, respectiv pixelilor din vecinătatea N_4 .

O altă variantă este utilizarea unor nuclee ce reprezintă aproximări ale suprafeței lui *Gauss*, date de funcția:

Exemple de nuclee de diferite dimensiuni și pentru valori diferite ale deviației σ sunt date în continuare:

Se observă ponderarea diferită a pixelilor din vecinătate în funcție de poziția lor față de pixelul central.

Trebuie remarcat faptul că nucleele prezentate mai sus pot fi aplicate consecutiv imaginii, însă o aplicare excesivă conduce la efectul denumit **blur** (ceață) și va determina pierderea detaliilor.

În figura 6 este prezentată o operație de filtrare a imaginii din figura 3 la care s-a adăugat un zgomot reprezentat de un număr de pixeli albi și negri generați aleator. A fost utilizat de două ori nucleul w_1 de dimensiune 3×3 . Ultima imagine prezintă efectul de "blur" prin aplicarea nucleului w_5 de patru ori succesiv pe jumătatea dreaptă a imaginii originale din figura 3.

