

UNIVERSITATEA POLITEHNICA BUCURESTI
FACULTATEA DE AUTOMATICA SI CALCULATOARE

ANALIZA SI EXTRAGEREA AUTOMATA A
CONTINUTULUI DOCUMENTELOR

**UN ALGORITM RAPID FOLOSIND MORFOLOGIE
PENTRU DETECTIA INCLINARII DOCUMENTELOR
SCANATE**

Iacob Diana
GMRV

Cuprins

Introducere	3
Metode existente	4
Algoritmul propus	6
Imbunatatirea algoritmului	10
Concluzii	18

Introducere

Deși mulți algoritmi de procesare a documentelor necesită un document drept, câteva grade de înclinare sunt inevitabile într-un document scaneat.

Algoritmii de procesare s-ar putea să nu fie capabili să producă rezultate satisfăcătoare pe un document oblic.

De obicei, este necesar un algoritm mai complex pentru a ajunge la rezultate satisfăcătoare, pentru simplul motiv că un document îndreptat permite o reprezentare mai compactă a componentelor sale, în particular a obiectelor text, cum ar fi cuvinte, linii și paragrafe.

Reprezentarea compactă duce la algoritmi mai simpli, mai eficienți și mai robusti.

Metode existente

Majoritatea algoritmilor pentru detectarea gradului de inclinare pot fi clasificați în 3 categorii :

(i) metoda proiectării profilului (projection profile method)

-documentul este proiectat cu diferite unghiuri. Apoi sunt identificate varfuri (datorate pozițiilor liniilor de text) și troughs (datorate spațiilor goale dintre linii). Unghiul care da diferența maximă între varfuri și troughs este acceptat ca fiind unghiul de inclinare.

Această metodă a fost îmbunătățită prin partitionarea verticală a imaginii într-un număr de fasii. Apoi profilul proiectiei orizontale este calculat pentru fiecare fasia, unde din corelația profilurilor fasiilor vecine, este estimat unghiul de inclinare. Metodele de proiectie a profilului sunt bune pentru a estima un unghi de inclinare de ± 10 grade.

(ii) împartirea în categorii a vecinilor cei mai apropiați (clustering of nearest neighbors)

- Pixelii sunt grupați în sectoare de-a lungul liniei mediene și a liniei de baza. Etichetarea componentelor este folosită pentru a calcula înălțimea medie a componentelor. Componentele cu un nr mic de pixeli (i, sau virgule) și caracterele înalte (ale căror înălțimi sunt mai mari decât media) sunt înlăturate. Acest lucru duce la identificarea destul de precisă a pixelilor aliniați de-a lungul liniei mediene. Inclinația documentului este o medie a valorilor de inclinare obținute pentru fiecare linie.

(iii) transformarea Hough

- Transformarea planului imaginii într-un spațiu de parametrii

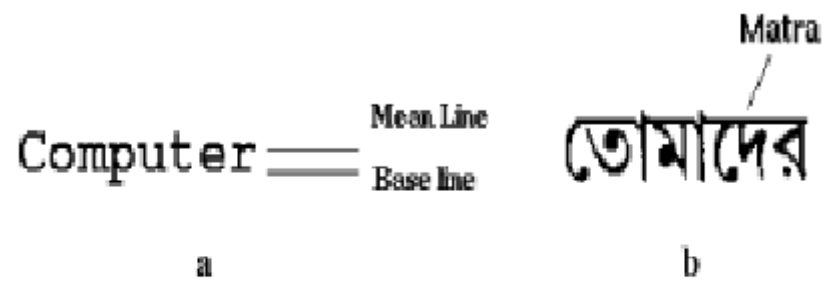


Fig. 1a. Mean line and base line b 'Matra' joining the characters of a 'Bangla' word

Algoritmul propus

Acest algoritm foloseste operatii matematice morfologice, i.e. dilatare, erodare, deschidere si inchidere pentru a raspandi (smear) liniile de text in benzi negre solide si pentru a inlatura denivelarile din benzi.

Pasi:

1. „Inchide” imaginea cu elementul de structurare a liniei ES (line structuring element SE) pentru a forma benzi negre solide corespondente fiecarei linii de text. Unele portiuni din imagine pot forma pete mari din cauza acestei operatii.
2. „Deschide” imaginea (inchisa) cu un ES patrat mic pentru a inlatura denivelarile formate datorita prezentei ascenders si descenders, etc. In consecinta, liniile si arcele de grafica sunt deasemenea inlaturate.
3. Scaneaza vertical imaginea „deschisa” pentru a inregistra toate tranzitiile de la 1 la 0, i. e., pixelii liniei de baza (base line) ai liniei de text. Pot aparea unele linii sau curbe mici datorita petelor aparute la pasul anterior.
4. Foloseste etichetarea pe componente pentru a selecta liniile ale caror lungimi sunt mai mari decat un prag. In plus, inlatura liniile care nu sunt drepte.
5. Identifica doua puncte in acele linii pentru a selecta unele portiuni de la inceput, respectiv sfarsit.
6. Gaseste inclanarea tuturor liniilor si declara valoarea mediana ca fiind inclinarea intregului document.

Se observa ca liniile generate la pasul 3, datorita prezentei elementelor grafice, nu sunt considerate la pasul 6 din cauza pasilor 4 si 5.

Algoritmul propus foloseste operatii morfologice de inchidere pentru a raspandi liniile de text in benzi negre folosind o linie ES cu lungime de l pixeli.

Benzile negre vor avea denivelari deasupra liniilor mediane si sub liniile de baza datorita prezentei ascenders si descenders la caractere.

Pentru a elimina acele denivelari, imaginea inchisa este deschisa cu un ES rectangular mic de dimensiune 5×5 . Aceasta alegere de ES pentru inchidere si deschidere a rezultate bune pentru fonturi comun utilizate ca 10pt-16-pt si pt o rezolutie de 300dpi.

Dupa inchidere si deschidere obtinem benzi negre de grosime (aprox) uniform separate cu spatii interliniare.

O tranzitie de la 1 la 0 scanata vertical este mapata ca o imagine noua. Aceasta da conturul portiunii de jos a fiecarei benzi negre (smeared text line) ca line elements (linii groase de 1 pixel).

Exista discontinuitate in linii datorita prezentei proiectiei portiunilor rectiliniare (datorita aparitiei multiplelor caractere cu descenders) in partea inferioara a benzilor, dar in medie obtinem linii orientate catre inclinarea textului.

Prezenta petelor in benzile negre rezultate cauzeaza pauze in liniile de baza ale textului. Acest lucru poate duce la segmente scurte si unghiuri de inclinare inexacte.

Portiunea de jos este selectata in mod deliberat dupa cum am vazut mai putine ondulatii in portiunea de jos din cauza urmatoarelor motive:

1. Numarul descendentilor dintr-un fisier text este relativ mai mic decat numarul ascenders. Procentajul aparitiei ascenders si descenders in fisierele text este studiat statistic examinand 50 de fisiere text de dimensiuni variate.
2. In multe scrieri (incl. Eng) caracterele sunt aliniate liniei de baza.

Se intentioneaza selectatea liniilor lungi deoarece dau estimari mai bune. Acest lucru este facut la pasul 4 prin etichetarea componentelor. Liniile care nu sunt drepte sunt filtrate si scoase. Pentru o acuratete si mai buna, intai, „prune” liniile pana la o lungime egala cu latimea ES de la inceputul si sfarsitul fiecarei linii, deoarece am vazut distorsiuni in acele parti datorate liniilor curbe din corpul caracterelor. Apoi se iau in calcul numai acele linii ale caror lungimi sunt mai mari decat un prag predefinit (de cel putin 4x latimea ES) si se calculeaza inclinarea fiecarei linii. In final se calculeaza inclinarea prin luarea valorii mediane calculate pentru liniile mari selectate si procesate la pasii 4 si 5 .

Table 1. % of characters with ascenders and descenders in English text

Ascenders		Descenders	
Single	Pairwise	Single	Pairwise
19.40	14.89	4.00	0.71

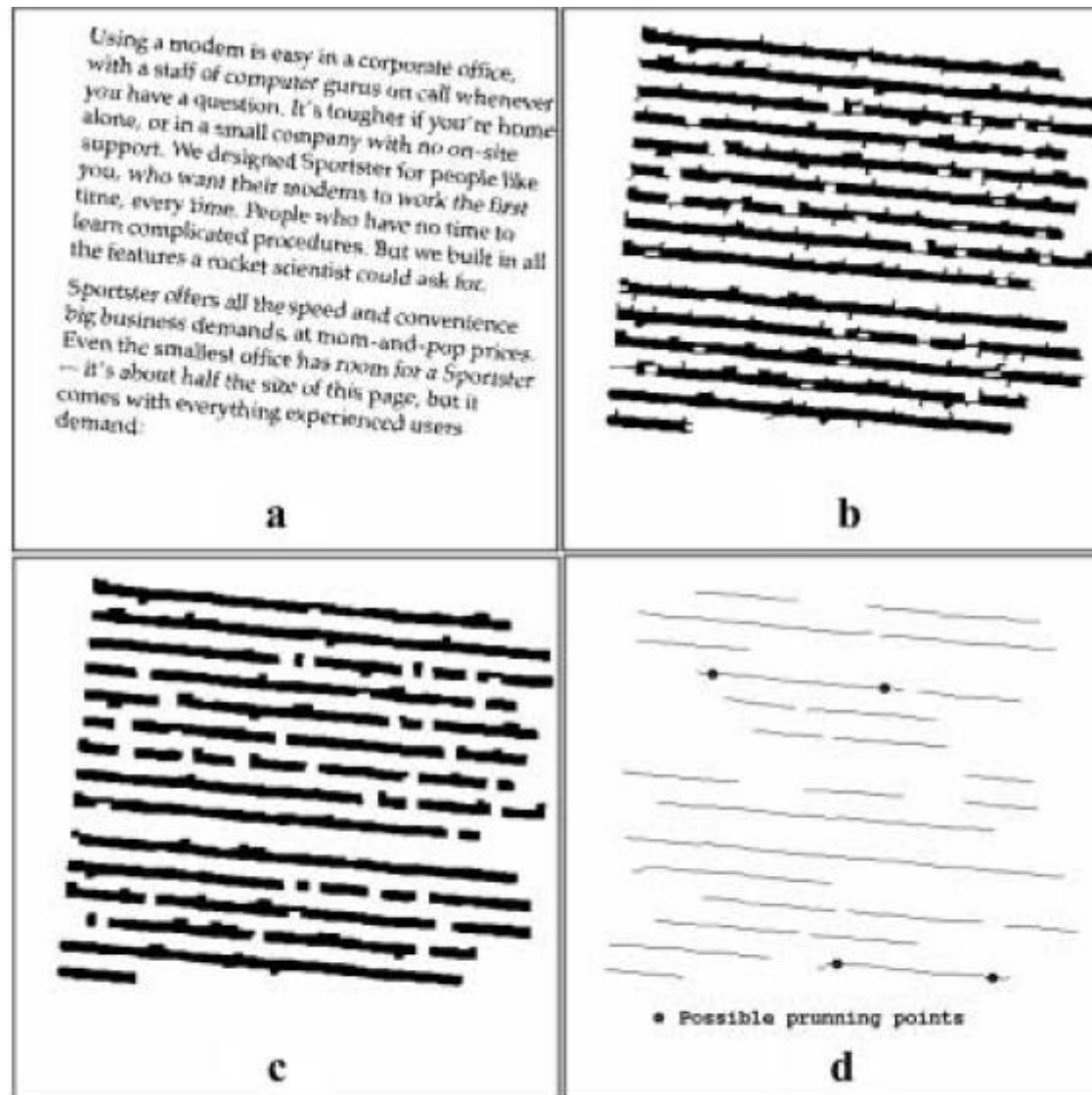


Fig. 2a. Original document in binary b closing original binary image with 4×12 SE c opening the closed image with 5×5 SE d after transition and rejection of small lines operation

Cresterea vitezei cu dilatarea si etichetarea componentelor

Timpul necesar pentru calculul inclinarii poate fi redus substantial prin utilizarea operatiunilor de dilatare si erodare in loc de inchidere, respectiv deschidere. Performanta nu va fi degradata in contextul algoritmului propus. De exemplu, scopul inchiderii aici este de a imprastia caracterele si cuvintele pentru a forma benzi negre. Dilatarea poate fi folosita in acest context din moment ce se obtine acelasi efect cu eleongarea componentelor pe directia orizontala. Cum mai tarziu inlaturam inceputul si sfarsitul fiecarei linii, dilatarea nu introduce eleongarea orizontala a pixelilor din capat indiferent de inclinare. Deshiderea poate fi inlocuita cu erodarea. Cresterea vitezei este obtinuta prin formularea unui modul special de dilatare potrivit elementului de structurare a liniei folosit pentru smearing. Metoda este acum descrisa.

Presupunem ca originea ES este la pixelul cel mai din stanga. Acum, in loc de dilatarea tuturor pixelilor din foreground in directia orizontala, aliniem originea ES cu un pixel foreground al imaginii de input, si toti pixelii de-a lungul ES liniar sunt schimbati in pixeli foreground in imaginea dilatata. Apoi, incepand de la cel mai din dreapta pixel al ES, traversam catre stanga si ne oprim la primul pixel foreground al imaginii input. Marcam acest pixel ca urmatorul pixel foreground care urmeaza sa fie dilatat daca nu este deja considerat ca unul, daca este, urmeaza ordinea scanarii raster pentru a gasi urmatorul pixel foreground care urmeaza sa fie dilatat. Aceasta abordare accelereaza intregul proces deoarece in text, aparitia pixelului foreground este frecventa si prin operatia prezentata mai sus, sunt uniti direct doi pixeli foreground situati la cele mai indepartate puncte din interiorul ES. Eroarea este duala dilatarii, asa ca este efectuata folosind acelasi modul si operatia complementara.

Pentru a obtine o mai buna crestere a vitezei, procesul de etichetare al componentelor este reformulat pt linii late de 1 pixel si conetate cate 8.

Pentru etichetarea componentelor, folosim o scanare verticala pentru a detecta cel mai din stanga pixel al unei componente neetichetate.

Apoi, portiunea ramasa a componenteii este etichetata dupa cum urmeaza: incepem cu cel mai din dreapta pixel pt ca prezenta lui are cea mai mare probabilitate.

Apoi, verificam pixelul din dreapta sus sau din dreapta jos. Observam ca este caz single scan 3neighborhood.

Procesul de etichetare al componentelor deasemenea stocheaza atat coordonatele punctelor de sfarsit ale fiecarei linii cat si punctele pana la care vom „prune” liniile.

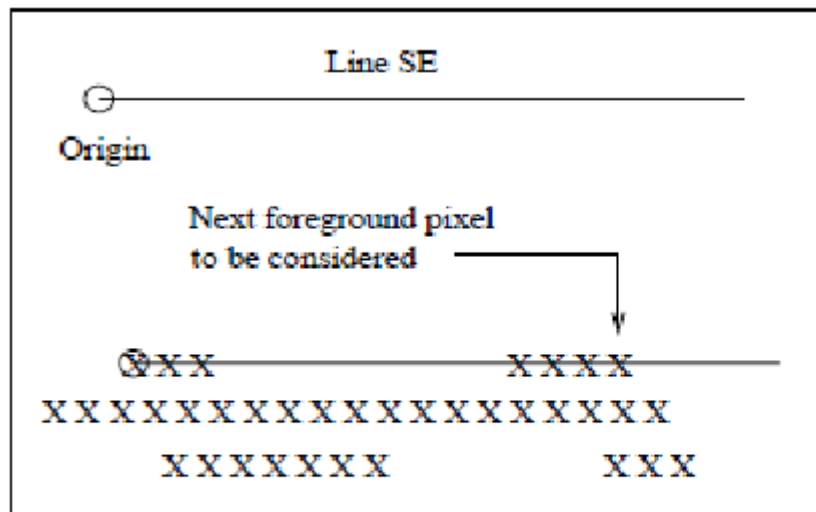


Fig. 3. Dilation with line SE to join two foreground pixels

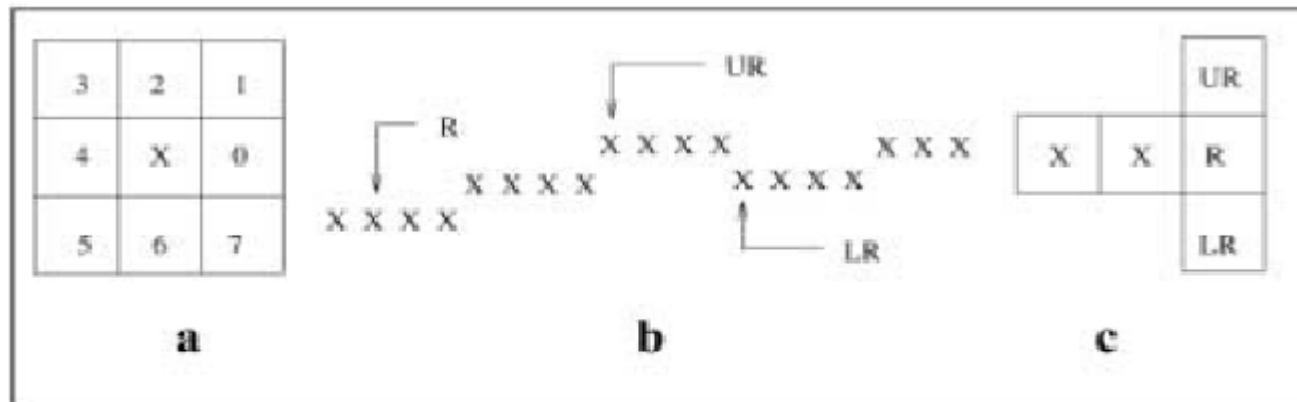


Fig. 4a. 8-neighboring pixels (0–7) around a foreground pixel 'x' **b** representation of a skewed line **c** 3 possible neighbors (R, UR, and LR) of a foreground pixel 'x' of a 1-pixel-wide line object

Table 2. Mean and standard deviation for five different methods (skew angle is given in degrees)

Skew angle	Mean	Mean	Mean	Mean	Mean
3	3.15	3.26	3.51	2.90	3.11
5	4.83	5.21	5.63	5.38	5.14
10	10.32	9.82	10.64	10.53	10.36
Skew angle	SD	SD	SD	SD	SD
3	0.193	0.342	0.562	0.498	0.318
5	0.219	0.390	0.301	0.517	0.212
10	0.118	0.218	0.827	0.631	0.229

আমাদের জীবন সিন্ধু মনোহর, সিন্ধু মনোহর মনোহর মনোহর মনোহর,
আমাদেরই জীবন সিন্ধু মনোহর, মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর,
আমাদের মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর,
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর,
যে মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর।
সিন্ধু মনোহর মনোহর মনোহর মনোহর মনোহর।

Fig. 5. A 512 × 512 handwritten image in 'Bangla' script

Table 4. Test results for a handwritten document in 'Bangla' script

Angle of rotation	Hough transform (A)	Morphology-based (E)
0	1.0	0.7
+5	6.0	4.92
-5	+6	+5.44
+10	11.0	10.93
-10	+9.0	+9.74

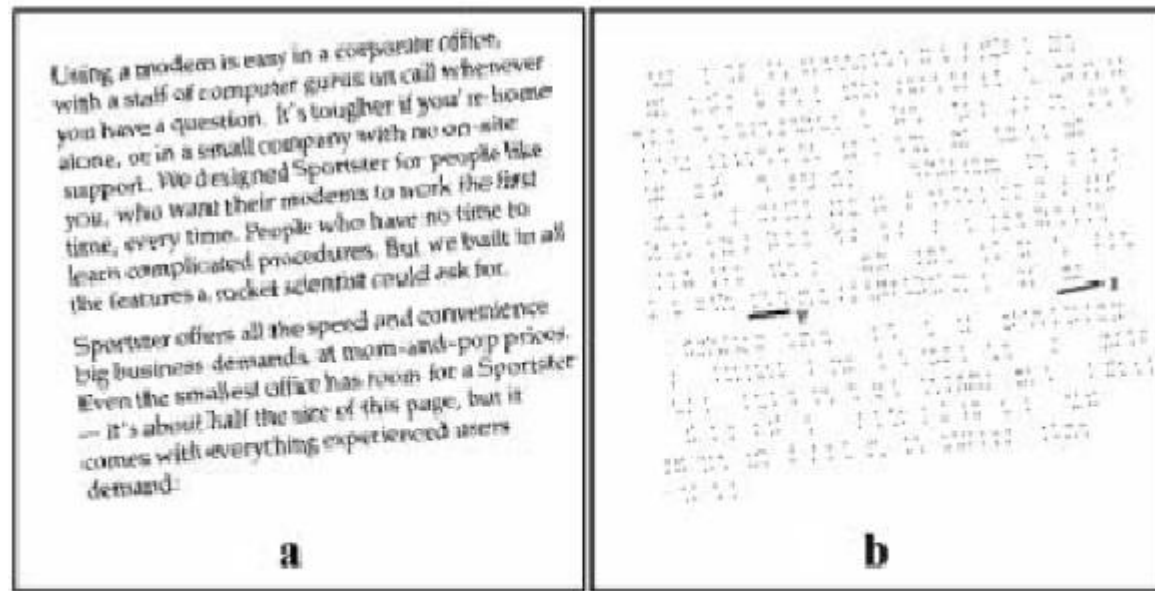


Fig. 6a. Distorted document image **b** top-left and bottom-right points of each component; two lines X and Y are drawn to show the points considered for forming 'ini-line' and a line with actual skew, respectively

Using a modem is easy in a corporate office, with a staff of computer gurus on call whenever you have a question. It's tougher if you're home alone, or in a small company with no on-site support. We designed Sportster for people like you, who want their modems to work the first time, every time. People who have no time to learn complicated procedures. But we built in all the features a rocket scientist could ask for.

Sportster offers all the speed and convenience big business demands, at mom-and-pop prices. Even the smallest office has room for a Sportster — it's about half the size of this page, but it comes with everything experienced users demand:

a

Using a modem is easy in a corporate office, with a staff of computer gurus on call whenever you have a question. It's tougher if you're home alone, or in a small company with no on-site support. We designed Sportster for people like you, who want their modems to work the first time, every time. People who have no time to learn complicated procedures. But we built in all the features a rocket scientist could ask for.

Sportster offers all the speed and convenience big business demands, at mom-and-pop prices. Even the smallest office has room for a Sportster — it's about half the size of this page, but it comes with everything experienced users demand:

b

Fig. 7a. Original document image **b** synthetically degraded image

Concluzii

Degradarea documentului nu are nici un efect asupra metodei propuse, deoarece textul este transformat in benzi negre, prin urmare toate discontinuitatile sunt umplute.

Acest algoritm poate manipula documente cu zgomot aditiv aleator , deoarece documentul este „deschis”cu un ES mic (pasul 2 din algoritm). Acest pas sterge efectiv pixelii din foreground.

Algoritmul functioneaza bine pentru unghiuri de inclinare mici (+-3 grade), deoarece depinde de lungimea ES, inclinare si spatiul dintre linii. Sansa formarii de poduri creste daca inclianrea este mare si spatiul dintre linii este mic.