



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale  
2007-2013



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

**AEACD**

**23. Document Clipping - Merging**

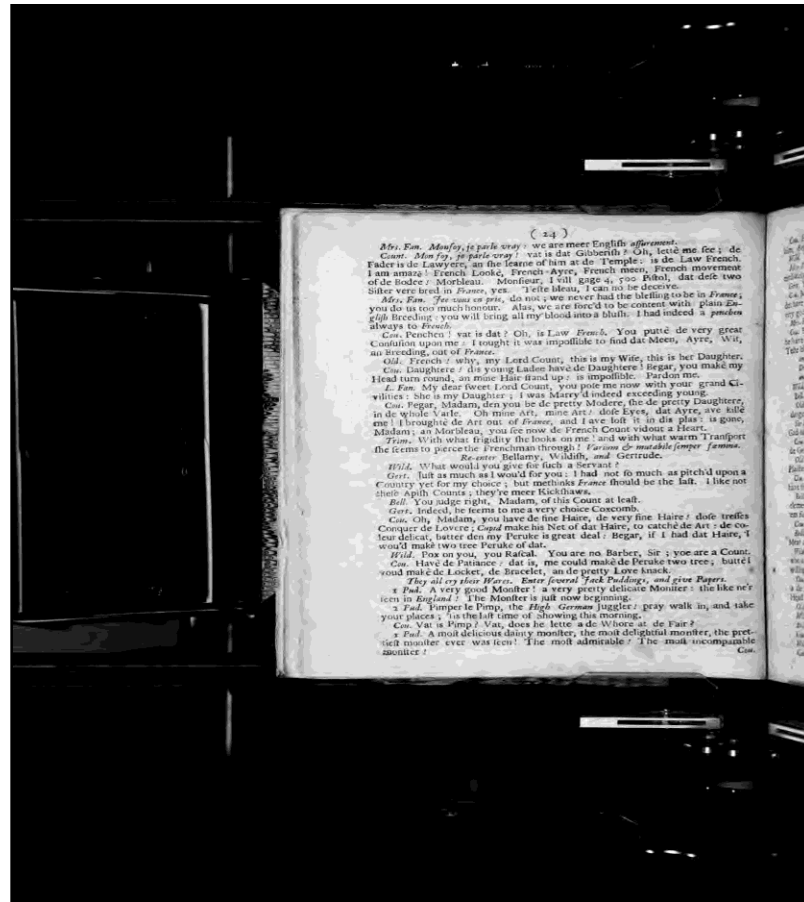
# Document Clipping

- Taierea documentului
- Se refera la decuparea partii importante din imaginea scanata (extragerea paginii de document si eliminarea backgroundului produs de scanner de exemplu)
- Simplifica procesarile ulterioare si reduce vizibil timpul de rulare al viitorilor algoritmi folositi in extragerea informatiei

# Eliminarea backgroundului

- Procesul de scanare depinde de aparatele folosite si nu va rezulta niciodata acelasi output
- De aceea este necesara taierea partii ne semnificative din imaginea scanata
- Aceasta reprezinta de cele mai multe ori backgroundul produs de scanner deoarece dimensiunea suprafetei de scanare este mai mare decat dimensiunea paginii de input

# Eliminarea backgroundului



( 24 )

*Abt. Fem. Monsieur, je parle voy : we are meer English affirmant.*  
*Count. Monsieur, je parle voy ! vas is dat Gibberish ? Oh, letie ma. Coe : de*  
*Faire is de Lawyere, an the Isaac of'm an de Temple : is de Law French,*  
*I am amaze ! French Look, French Ayre, French rassu, French mouvement*  
*of de booke : Monsieur ! I fill gage a, you Prol, dat dret two*  
*Sister very head in France, yes : l'esse bleau, I can no be deceiv.*  
*Abt. Fem. Je vas en pris, do not, we never had the blessing to be in France,*  
*you do us too much honour. Alas, we are forc'd to be content with plain En-*  
*gish Breeding : you will bring all my blood into a bluth. I had indeed a penchen*  
*aboyce to French.*  
*Count. Penchen ! vas is dat ? Oh, is Law French. You putie de very great*  
*Confusion upon me : I thought it was impossible to find dat Meev, Ayre, Wit,*  
*an Spreeding, out of France.*  
*Abt. French : why, my Lord Count, this is my Wife, this is her Daughter.*  
*Count. Daughter ! its young Lady have de Daughter ! Begar, you make my*  
*Head turn round, an mine Hair stand up : is impossible. Pardon me.*  
*Count. My dear sweet Lord Count, you putie me now with your grand Ci-*  
*vilities : she is my Daughter : I was Marry'd indeed exceeding young.*  
*Count. Begar, Madam, den you be de pretty Modere, be de pretty Daughter,*  
*in de whole Vaire. Oh mine Art, mine Art : dole Eyes, dat Ayre, ave kind*  
*me ! I thought de Art out of France, and I ave lost it in dis place : is gone,*  
*Madam ! an Monsieur, you see now de French Count vidout a Heart.*  
*Tim. With what dignity she looks on me ! and with what warm Transport*  
*she seems to perceive the Frenchman through ! *Parsons de-mutabile semper famula.**  
*Re-entr. Bellamy, Whilist, and Gertrude.*  
*Will. What would you give for such a Servant ?*  
*Gertr. Just as much as I would for you : I had not so much as pitch'd upon a*  
*Country yet for my choice : but methinks France should be the last. I like not*  
*these Apth Counts : they're meer Richshaw.*  
*Will. You judge right, Madam, of this Count at least.*  
*Gertr. Indeed, he seems to me a very choice Coxcomb.*  
*Count. Oh, Madam, you have de fine Haire, de very fine Haire ! dole tresita*  
*Compter de Lovere, Casid make his Net of dis Haire, to catch de Art : de co-*  
*leur delicat, latter den my Feruke is great deal ! Begar, if I had dat Haire, I*  
*would make two tree Feruke of dat.*  
*Will. Pox on you, you Katsch. You are no Barber, Sir : you are a Count.*  
*Count. Have de patience : dat is, me could make de Feruke two tree, but I*  
*would make de Locket, de Breelet, an de pretty Love Knack.*  
*They all cry Alas Hère. Enter several Jack-Pudding, and give Pappes.*  
*Paul. A very good Monster ! a very pretty delicate Monster : she like ney*  
*fers in England ! The Monster is just now beginning.*  
*Paul. Pumper le Pomp, the High German Juggler : pray walk in, and take*  
*your places, 'tis the last time of showing this morning.*  
*Count. Vat is Pamp ! Vat, does he lette a de Whore at de Fair ?*  
*Paul. A most delicious dainty monster, the most delightful monster, the pret-*  
*tiest monster, ever was seen ! The most admittable ! The most incomparable*  
*monster !*  
*Count.*

# Eliminarea backgroundului

- Pentru a elimina backgroundul se foloseste un algoritm foarte simplu
- Se calculeaza valoarea grayscale a fiecarului pixel din imagine si astfel se calculeaza suma valoriiilor grayscale pe fiecare linie si coloana
- Toate liniile ce au suma grayscale mai mica ca 60% din valoare sumei maxime grayscale pe linii sunt eliminate
- Toate coloanele ce au suma grayscale mai mica ca 30% din valoare sumei maxime grayscale pe coloane sunt eliminate

# Extragerea paginii

- Prin procesul anterior se va elimina backgroundul negru produs de scanner
- Acest pas nu este suficient deoarece se dorește extragerea paginii, iar imaginea poate conține de asemenea elemente ale altei pagini (în cazul cartiilor) după cum s-a observat în imaginea exemplu
- Este necesar un alt pas pentru extragerea paginii

# Extragerea paginii

- Se calculeaza un scor pentru fiecare linie si coloana din imaginea rezultata dupa eliminarea backgroundului produs de scanner
- Scorul este acelasi ca si inainte: suma grayscale; doar ca de aceasta data pentru fiecare pixel cu o valoare mai mare ca 100 (aproximativ valoarea de background) se adauga 200 pentru a scoate in evidenta diferenta dintre background si foreground

# Limitele orizontale

- Limitele orizontale ale paginii se gasesc in felul urmator:
  - Se parcurge imaginea de sus in jos si in momentul in care se gaseste o linie cu scorul mai mare ca 30% din valoare scorului maxim al liniilor procesul se opreste si linia curenta este aleasa ca limita superioara
  - Pentru limita inferioara algoritmul este identic doar ca se porneste de jos in sus



# Limitele verticale

- Procedeu este același ca și în cazul limitelor orizontale cu două mențiuni
- Se verifică coloanele de la stânga la dreapta și de la dreapta la stânga, iar valoarea de decizie este de 90% din valoarea maximă
- Algoritmul se aplică doar pe 10% din lățimea imaginii în stânga și 5% în dreapta în cazul paginilor impare și invers în cazul paginilor pare
- Scopul acestei limitări este îmbunătățirea timpului de rulare

# Rezultatul final

( 24 )

*Mrs. Fan.* *Monfey, je parle vray* : we are meer English *affurement*.

*Count.* *Mon foy, je parle vray* ! vat is dat Gibberish ? Oh, lettè me see ; de Fader is de Lawyere, an she learne of him at de Temple : is de Law French. I am amazè ! French Lookè, French Ayre, French meen, French movement of de Bodce ! Morbleau. Monsieur, I vill gage 4, 500 Pistol, dat dese two Silter vere bred in France, yes. Teste bleau, I can no be deceive.

*Mrs. Fan.* *Fee vous en prie*, do not ; we never had the blessing to be in France ; you do us too much honour. Alas, we are forc'd to be content with plain English Breeding : you will bring all my blood into a blush. I had indeed a *penchen* always to French.

*Con.* Penchen ! vat is dat ? Oh, is Law French. You puttè de very great Confusion upon me : I tought it was impossible to find dat Meen, Ayre, Wir, an Breeding, out of France.

*Old.* French ! why, my Lord Count, this is my Wife, this is her Daughter.

*Con.* Daughtere ! dis young Ladee havè de Daughtere ! Begar, you makè my Head turn round, an mine Hair stand up : is impossible. Pardon me.

*L. Fan.* My dear sweet Lord Count, you pote me now with your grand Civilities : She is my Daughter ; I was Marry'd indeed exceeding young.

*Con.* Pegar, Madam, den you be de pretty Modere, she de pretty Daughtere, in de whole Varle. Oh mine Art, mine Art ! dose Eyes, dat Ayre, ave killè me ! I broughtè de Art out of France, and I ave lost it in dis plas : is gone, Madam ; an Morbleau, you see now de French Count vidout a Heart.

*Trim.* With what frigidty she looks on me ! and with what warm Transport she seems to pierce the Frenchman through ! *Varium & mutabile semper fœmina.*

*Re-enter Bellamy, Wildish, and Gertrude.*

*Wild.* What would you give for such a Servant ?

*Ger.* Just as much as I wou'd for you : I had not so much as pitch'd upon a Country yet for my choice ; but methinks France should be the last. I like not these Apish Counts ; they're meer Kickthaws.

*Bell.* You judge right, Madam, of this Count at least.

*Ger.* Indeed, he seems to me a very choice Coxcomb.

*Con.* Oh, Madam, you have de fine Haire, de very fine Haire ! dose tresses Conquer de Lovere ; *Cupid* make his Net of dat Haire, to catchè de Art : de couleur delicat, batter den my Peruke is great deal : Begar, if I had dat Haire, I wou'd makè two tree Peruke of dat.

*Wild.* Pox on you, you Rascal. You are no Barber, Sir ; yoe are a Count.

*Con.* Havè de Patiance : dat is, me could makè de Peruke two tree ; buttè I sould makè de Locket, de Bracelet, an de pretty Love knack.

*They all cry their Wares. Enter several Jack Puddings, and give Papers.*

*1 Pad.* A very good Monster ! a very pretty delicate Moniter : the like ne'r seen in England ! The Monster is just now beginning.

*2 Pad.* Pimper le Pimp, the High German juggler ! pray walk in, and take your places ; 'tis the last time of Showing this morning.

*Con.* Vat is Pimp ? Vat, does he lette a de Whore at de Fair ?

*3 Pad.* A most delicious dainty monster, the most delightful monster, the prettiest monster ever was seen ! The most admirable ! The most incomparable monster !

*Con.*

# Document Clipping

- O alta aplicatie a document clipping este extragerea unei parti din document, unui contur poligonal (de exemplu al unei poze)
- Algoritmul se numeste “Beta-Star Shape” si este o imbunatatirea a algoritmul “Alpha-Shape” ce folosea ca distanta raza unui cerc, obtinandu-se astfel rezultate eronate in anumite cazuri (forma poligonala era dependenta de forma sferica si astfel putea sa omita elemente dorite sau sa includa elemente nedorite)

# Beta-Star Shape

- Inputul algoritmului îl reprezintă înfășurătoare convexă a elementului detectat în interiorul formei pe care dorim să o obținem
- Toate punctele din afara acestui poligon sunt puncte candidat, puncte ce pot fi incluse în poligon în următorii pași
- În algoritm se folosește un parametru  $d$ , ce va reprezenta distanța de unire, similar cu raza cercului din algoritmul “Alpha-Shape”

# Beta-Star Shape

- Daca doua puncte adiacente se afla la o distanta mai mica ca  $d$ , atunci segmentul ce le uneste este inclus in poligon
- Altfel procesul de rafinari succesive va imparti segmentul pentru a obtine cel mai bun candidat pentru incluziunea in forma “Beta-Star”
- Cel mai bun candidat este cel mai apropiat punct de margine ce nu va genera intersectii prin incluziune

# Beta-Star Shape

- Daca acest punct nu se gaseste pe segment se vor evalua ceilalti candidati
- Acest proces de descoperirea a candidatului se realizeaza pentru fiecare muchie a poligonului, rezultand includerea acestora in noua forma
- Algoritmul se repeta pentru noile muchii pana cand nu mai exista candidatii ce nu ar genera intersectii sau pana cand toate muchiile au dimensiuni mai mici ca  $d$

# Beta-Star Shape


**all**

**BLUE DEVILS**

## ftritt, Mr. Moore



Er in einem neuen Team – und "Shutdown-Tony" auf Partnersuche ...  
 (US-Boys, 1995)  
 ga-Karriere-  
 in blieb  
 geht der  
 im Dei-

nicht überlegt es  
 seine US-Boys  
 in anders, denn  
 wie "Tony" ist  
 wirklich sehr und  
 sein Name immer  
 rückt an.  
 Außerdem hat  
 er seine Liebe zu  
 mit strecken-  
 genick war hoch  
 zu gut und ich  
 ich ein bisschen  
 ge hätte in der  
 ja spielen kann,  
 eine Sorge, Tony  
 Partie morgen  
 ich eine (im Be-  
 im Cafe Gestern  
 ist da kann man  
 Kontakte immer  
 in Augenblicke  
 Kontaktanbahn-  
 dem Spiel wird  
 er höher seine  
 haben. Wie es  
 bei den Partien  
 immer wieder ist  
 schwer, ein-  
 e. „Wir haben  
 ein relativ klei-  
 n, aber das ist  
 eben. Die Fans  
 ich ein toller  
 er“, erklärt der  
 aber auch keine  
 r Probleme. Bis




lang können die  
 Duathlon erst zwei  
 Stundenlangspiele  
 (beide gingen ser-  
 ious) absolvieren,  
 die Bergtouren  
 gegen Köln und  
 Düsseldorf haben  
 dagegen bereits  
 ein. Das ist heute  
 für ihn, denn da  
 durch lebt er in  
 täglich Wettkamp-  
 fern. Der könnte  
 sich gegen ein  
 Team wie die Blue  
 Devils stellen zu-  
 schen. „Fürchtet  
 Moore, der sich  
 schon auf das Spiel  
 gegen die Ravens  
 freut. „Es ist  
 nur schade, dass die  
 Partie nicht im Volk-  
 parkstadion stattfin-  
 den kann, dort habe  
 ich immer besonders  
 gute gespielt. Die At-  
 mosphäre war etwas  
 ganz besonderes.“

**NACHMITTAGS-TIME** – Tony Moore (links) re-  
 dert bei Birkhof Schulen. Mit einigen ameri-  
 kanischen Leichtathleten stand der Ex-Football-  
 Reda und Antwort zum Thema US-Boys zu-  
 einem eine deutsche Frau.

**AUF DER SUCHE NACH DEM LEDER-ELI**

**Hamburg-Tour für Football-Fans – oder wie ich diesmal zum Stadion finde ...**



# Aplicatii

- Incadrarea elementelor descoperite in document astfel incat sa nu se realizeze o suprapunere a acestora
- In functie de caz se pot folosi dreptunghiul de incadrare, infasuratoarea convexa sau algoritmul prezentat anterior “Beta-Star Shape”