BOBCAT Review

July 10, 2008

David Doermann
Wontaek Seo
Elena Zotkina
Mudit Agrawal

Agenda

- Review of Goals
- Progress on:
  - Datasets
  - Evaluation Methodology
  - Segmentation Survey and Tools
- Open Discussion of Additional Plans

Overview of Goals

- Transition the test methods, metrics, and procedures … as part of the assessment infrastructure,
- Provide tools … to extend groundtruthed datasets to include Arabic Anfal images.
- Provide test designs, data analysis procedures, and interpretation guidelines for evaluating COTS and GOTS OCR systems and other DIP tools

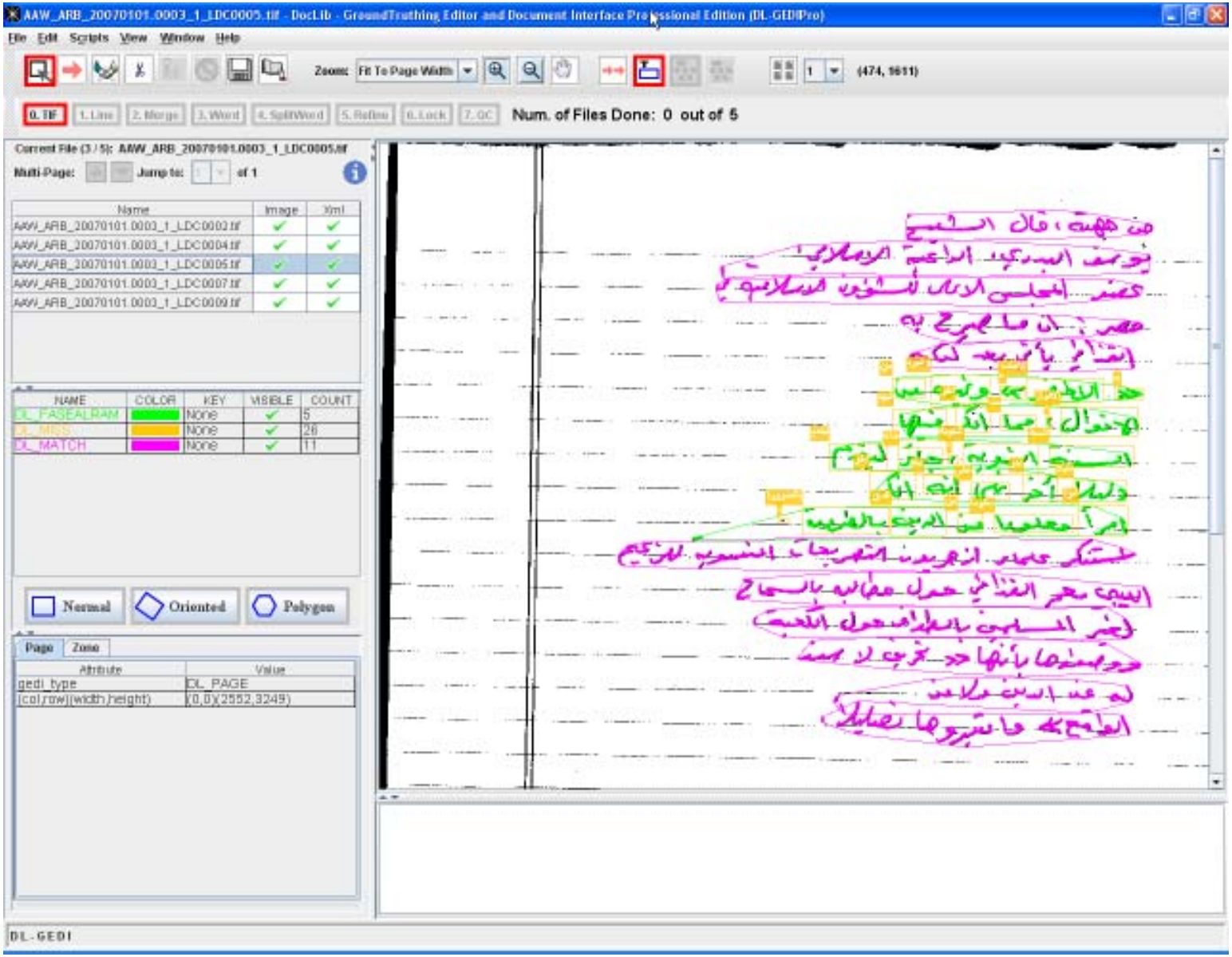
- Provide a basis for Phase II of MadCat
  - Groundtruthing Guidelines
  - Evaluation Metrics
  - Data Representations
- Issues:
  - How do we extend representations to Handwriting
  - How do we represent uncertainty
  - How do we provide a dataset useful for various tasks
    - segmentation, OCR, content labeling, etc

Specific Tasks

- Data
  - Zone Classification and Segmentation GT
- Tools
  - Update GEDI to allow handwritten data rep
- Evaluation
  - Zone Classification Tools

GEDI Tool

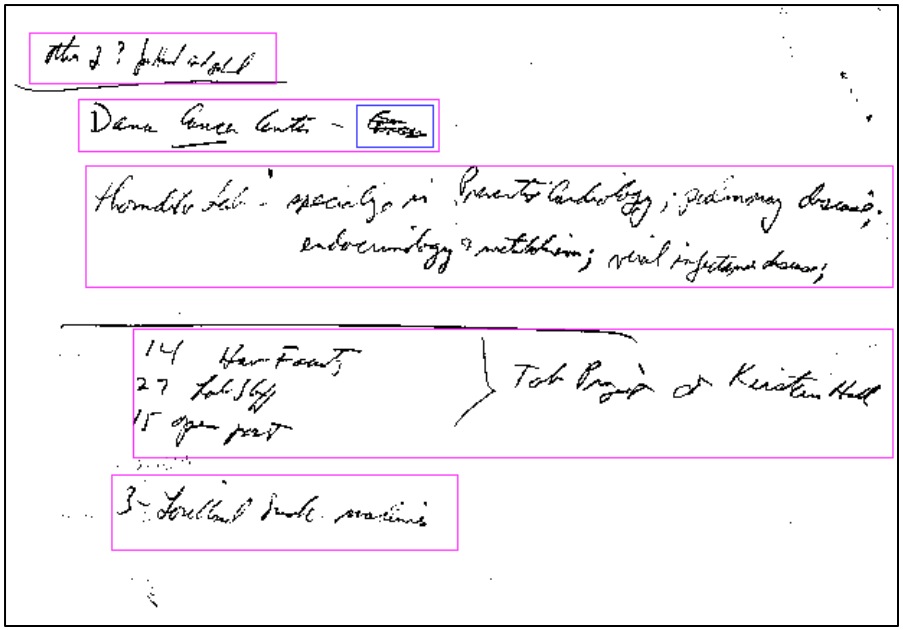
- Overview
  - Generic Tool for Representing Regions and Attributes on images
- Project Specific Extensions
  - Reading Order
  - Representation of Run Length Encoded Data for Line Segmentation
  - Direct Integration of Evaluation Capabilities



U.S. Army Research Laboratory

Data Sets

- Segmentation/Classification
  - 26,007 pages of Tobacco Litigation Corpus
  - 320,000+ zones
  - Useful for Large Evaluations



Statistics

Category	Documents
advertisement	451
bibliography	158
calendar	44
drawings	597
email	962
fax	815
foreign	761
form	1,407
graphic	518
handwritten	2,766
letter	2,561
list	395
marginalia	888
memo	1,893
newspaper	615
periodical	22
photograph	227
questionnaire	188
report	985
tables	690
Total Documents	**16,943**
Page Count	**26,007**

Zone Type	Count
FORM	3,679
GRAPHICS	3,430
HANDPRINT	50,138
Image	1,484
LOGO	4,070
MACHINEPRINT	210,696
MARKUP	27,533
SIGNATURE	5,552
STAMP	5,074
TABLE	5,559
TITLE	5,800
Total	**323,015**

Anfal Data

- Line of text GT with polygons
- Lines Split by
  - Physical Location
  - Change in Attribute – hand/machine, size
- Reading Order used to link segments of a line

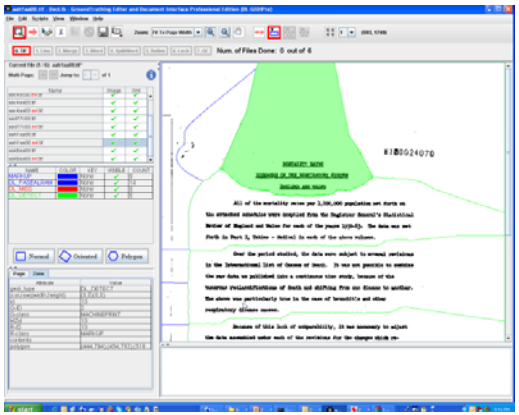
MADCAT

- Set of Word Boxes Mapped to Lines
- Run Length Encoded Data in each zone

- Algorithms return Polygons which are matched at the line level.

Agenda

- Review of Goals
- Progress on:
  - Datasets
  - Evaluation Methodology
  - Segmentation Survey and Tools
- Open Discussion of Additional Plans



University of Maryland

U.S. Army Research Laboratory

Evaluation Methodology and Software

Wontaek Seo

David Doermann

Evaluation Modules

- Zone Classification
- Segmentation
  - Line Segmentation
  - Zone Segmentation

General Concept

- Given two zones to be compared, calculate the matching score if there is at least one shared ON pixel

- Four types of result
  - MATCHED: location and zone type
  - DETECTED: location but not zone type
  - FALSE: Extra zone in Results
  - MISSED: Zone not matched from GT

- Threshold is set to determine which zones are matched for “detection”
- Zone types “can” be used for matching
- Software is integrated into DocLib
- Full match matrix is built to store the score of each pair of zones.

UNIVERSITY OF 18 56 MARYLAND
U.S. Army Research Laboratory

Matching score

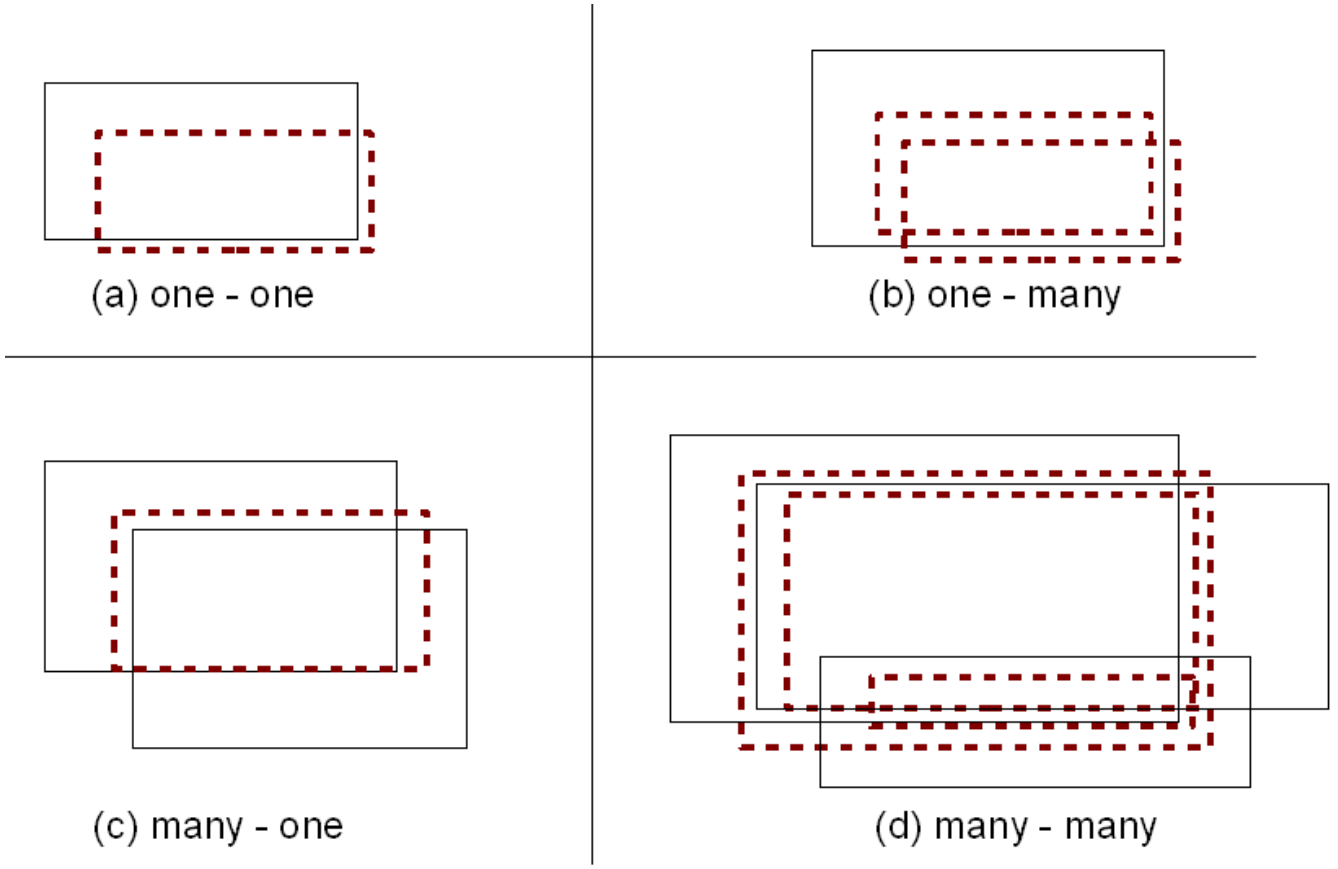
- I = set of all ON pixel in Image
- R_i = set of all ON pixel in the result zone
- G_j = set of all ON pixel in the ground truth zone
- *T(s)* = function that count the elements of set s

$$MatchScore(i, j) = \frac{T(G_j \cap R_i \cap I)}{T((G_j \cap R_i) \cap I)} \times 100$$

Types of result

- MATCHED
  - MatchScore(i,j) ≥ threshold
  - L(i) = L(j)
- DETECTED
  - MatchScore(i,j) ≥ threshold
  - L(i) ≠ L(j)
- FALSE
  - MatchScore(i,all) < threshold
- MISSED
  - MatchScore(all,j) < threshold

Matching examples



one-one

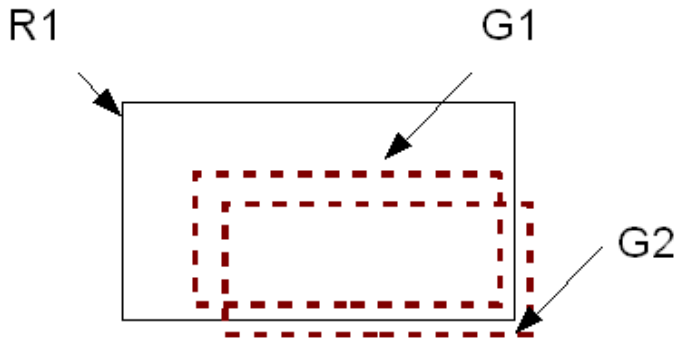
(a) example

result\GT	G1
R1	85.00%

(b) matching score

- Representation
  - L(A) : Label of A
- L(R1) = L(G1)
  - R1 is matching to G1
- L(R1) ≠ L(G1)
  - R1 is detecting G1 w/ the different label

one-many



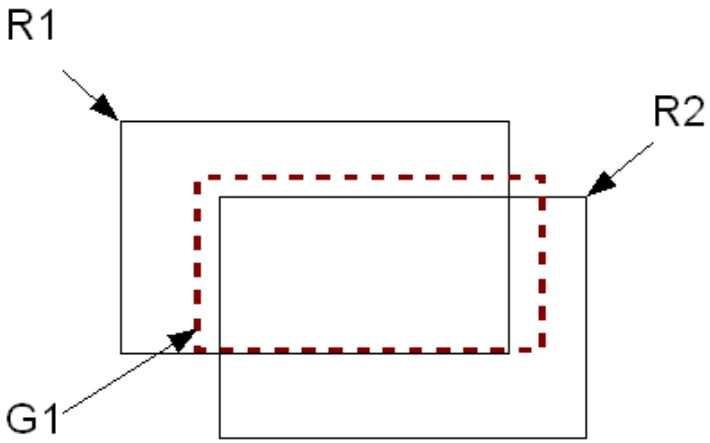
(a) example

result\GT	G1	G2
R1	90.00%	85.00%

(b) matching score

- L(R1) = L(G1)=L(G2)
  - compare the matching scores
    - R1 is matching to G1
    - G2 is missing
- L(R1) = L(G2) ≠ L(G1)
  - R1 is matching to G2
  - G1 is missing
- L(R1) ≠ L(G1) ≠ L(G2)
  - compare the matching scores
    - R1 is detecting G1 w/ the different label
    - G2 is missing

many-one



(a) example

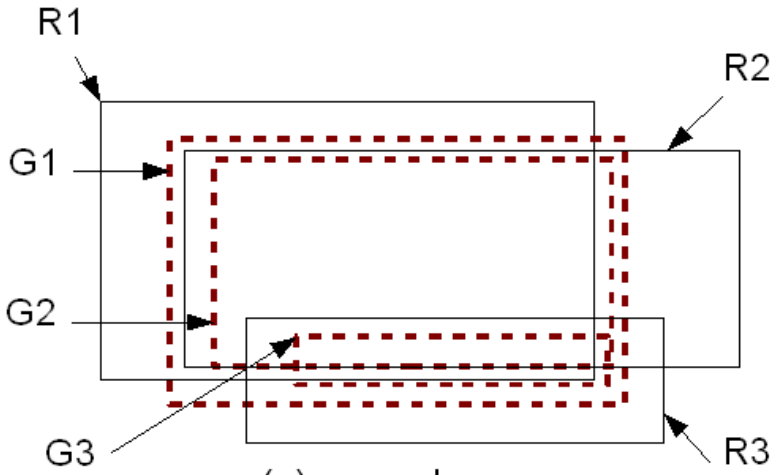
result\GT	G1
R1	95.00%
R2	90.00%

(b) matching score

- L(R1)=L(R2)=L(G1)
  - compare the matching scores
    - R1 is matching to G1
    - R2 is false alarm
- L(R1)≠L(R2)=L(G1)
  - R1 is false alarm
  - R2 is matching to G1
- L(R1),L(R2)≠L(G1)
  - compare the matching scores
    - R1 is detecting G1 w/ the different label
    - R2 is false alarm

UNIVERSITY OF 18 56 MARYLAND
U.S. Army Research Laboratory

many-many



(a) example

result\GT	G1	G2	G3
R1	90.00%	85.00%	50.00%
R2	80.00%	82.00%	45.00%
R3	30.00%	0.00%	85.00%

(b) matching score

- 1st step
  - find the set of matched zone which is not matched to same ground truth zone
- 2nd step
  - find the set of detected zone which is not matched in the 1st step
- The R which is not set at any steps is false alarm
- The G which is not set by any R is missing

Software

- PEZS : Performance Evaluation tool of Zone Segmentation
- Usage

  PEZS -r { FILE | DIR } -g { FILE | DIR } -img { FILE | DIR }
  [-o FILE -v DIR -m FILE -t NUM -detail -lid -rle -seg]

Note: Currently zone labeling eval is in Java… All will be in DocLib for final release.

Options

- **r { FILE | DIR }** : path to the result file or directory
- **g { FILE | DIR }** : path to the ground truth file or directory
- **img { FILE | DIR }** : path to the image file or directory
- **o FILE** : set file name of file to be saved
- **v DIR** : set directory where the GEDI type xml output for visualization will be saved
- **t NUM** : set the threshold of matching score

Options

- **m FILE :** result zones which is in a ground truth zone will be merged if it's type is in the FILE
- **detail** : result of each zone will be added to the output when it is set
- **rle** : run-length code will be added to the visualization output
- **seg** : label matching will not be performed when it is set

Software Output

```
Zone Segmentation Evaluation Result.
Generated on Sat Jul  5 11:10:57 2008

=========================
Result of Individual File
=========================

[O] : Detected, [-] : Detected w/ Different Type, [X] : Undet

AAW_ARB_20070101.0003_1_LDC0002.tif
===================================

Page ID : 1
-------------
[O]      1,   DL_TEXTLINEGT,   z10,   DL_TEXTLINEGT, 85.32%
[O]      2,   DL_TEXTLINEGT,   z11,   DL_TEXTLINEGT, 86.36%
[O]      3,   DL_TEXTLINEGT,    z2,   DL_TEXTLINEGT, 85.90%
[O]      4,   DL_TEXTLINEGT,   z12,   DL_TEXTLINEGT, 80.03%
[O]      5,   DL_TEXTLINEGT,    z1,   DL_TEXTLINEGT, 85.36%
[X]      6,   DL_TEXTLINEGT
[O]      7,   DL_TEXTLINEGT,   z13,   DL_TEXTLINEGT, 85.38%
[X]      8,   DL_TEXTLINEGT
[X]      9,   DL_TEXTLINEGT
[X]     10,   DL_TEXTLINEGT
[X]     11,   DL_TEXTLINEGT
[X]     12,   DL_TEXTLINEGT
[O]     13,   DL_TEXTLINEGT,    z4,   DL_TEXTLINEGT, 86.00%
[O]     14,   DL_TEXTLINEGT,    z0,   DL_TEXTLINEGT, 84.70%
[O]     15,   DL_TEXTLINEGT,   z14,   DL_TEXTLINEGT, 85.99%
[X]     16,   DL_TEXTLINEGT
[X]     17,   DL_TEXTLINEGT
[OVERALL] 9/0/8/17, 52.94%

AAW_ARB_20070101.0003_1_LDC0004.tif
===================================
```

```
==================
Summary of Results
==================

- Total number of R-Zone : 22033
- Accuracy of Zone Detecting : 31.19%

01. Information on Zones
========================

Label   Class of Zone        Number of Zone       Accuracy
-----------------------------------------------------------
1       DL_TEXTLINEGT                 22033         31.19%

02. Confusion Matrix
====================

 Result\GT      unmatch             1
-------------------------------------
   unmatch     0( 0.0%)* 12778(65.0%)
         1 15161(68.8%)    6872(31.2%)*

03. Result Table
================

Label   Total  Detected   Correct  Precsion    Recall   F-Score   Missing  FalseAlram
-------------------------------------------------------------------------------------
    1   19650     22033      6872    31.19%    34.97%    32.97%    65.03%      68.81%
```

Zone Classification

```
==================
Summary of Results
==================

- Total Number of Sample :  21786
- Overall Accuracy : 95.78%
- Average of Each Class Accuracy : 55.31%

01. Information on Classes
==========================
```

Label	Name of Class	Number of Sample	Accuracy
00	text_sm	20617	97.34%
01	ruling	201	61.69%
02	drawing	299	88.29%
03	table	76	46.05%
04	text_lg	51	64.71%
05	math	301	60.47%
06	halftone	144	83.33%
07	logo	13	0.00%
08	chm_drawing	80	51.25%
09	map	4	0.00%

02. Confusion Matrix

Out\GT	00	01	02	03	04
00	20068(97.3%)*	70(34.8%)	11(3.7%)	14(18.4%)	12(23.5%)
01	69(0.3%)	124(61.7%)*	0(0.0%)	1(1.3%)	1(2.0%)
02	93(0.5%)	1(0.5%)	264(88.3%)*	23(30.3%)	4(7.8%)
03	46(0.2%)	0(0.0%)	5(1.7%)	35(46.1%)*	0(0.0%)
04	19(0.1%)	1(0.5%)	0(0.0%)	0(0.0%)	33(64.7%)*
05	284(1.4%)	2(1.0%)	8(2.7%)	2(2.6%)	1(2.0%)
06	38(0.2%)	3(1.5%)	6(2.0%)	0(0.0%)	0(0.0%)
07	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)
08	0(0.0%)	0(0.0%)	5(1.7%)	1(1.3%)	0(0.0%)
09	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)

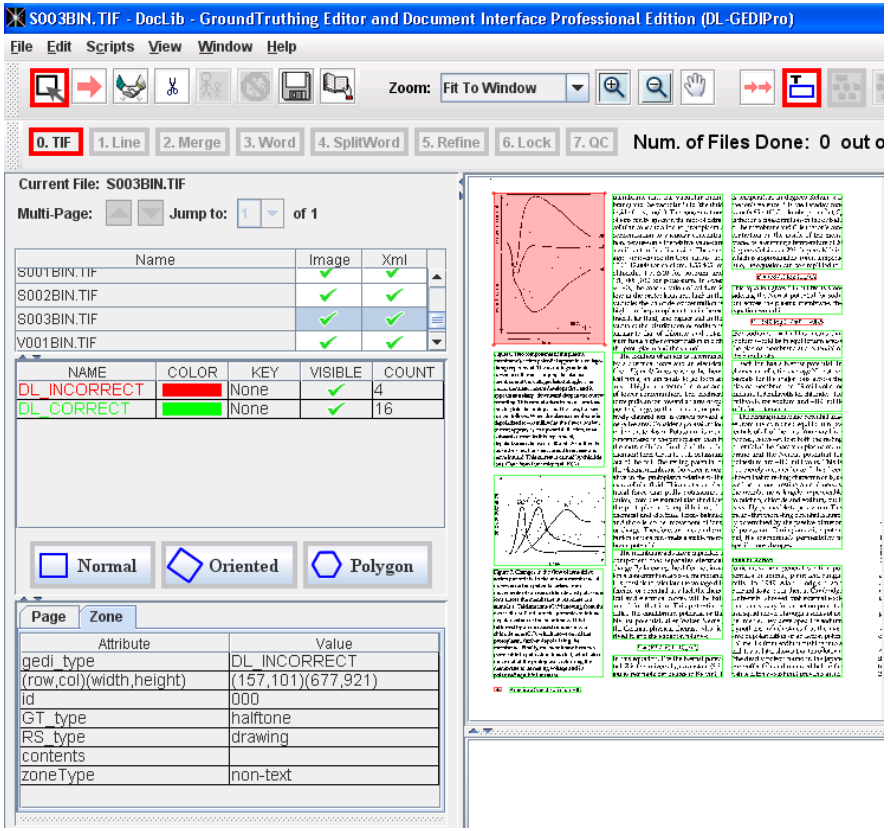
05	06	07	08	09
106(35.2%)	5(3.5%)	7(53.8%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	1(7.7%)	0(0.0%)	0(0.0%)
9(3.0%)	18(12.5%)	0(0.0%)	9(11.3%)	4(100%)
0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	4(30.8%)	0(0.0%)	0(0.0%)
182(60.5%)*	0(0.0%)	0(0.0%)	30(37.5%)	0(0.0%)
0(0.0%)	120(83.3%)*	0(0.0%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	0(0.0%)*	0(0.0%)	0(0.0%)
4(1.3%)	1(0.7%)	1(7.7%)	41(51.2%)*	0(0.0%)
0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)*

03. Precision and Recall

Class\Eval	precision	recall	detected	correct	total
00	98.89%	97.34%	20293	20068	20617
01	63.27%	61.69%	196	124	201
02	62.12%	88.29%	425	264	299
03	40.70%	46.05%	86	35	76
04	57.89%	64.71%	57	33	51
05	35.76%	60.47%	509	182	301
06	71.86%	83.33%	167	120	144
07	0.00%	0.00%	0	0	13
08	77.36%	51.25%	53	41	80
09	0.00%	0.00%	0	0	4

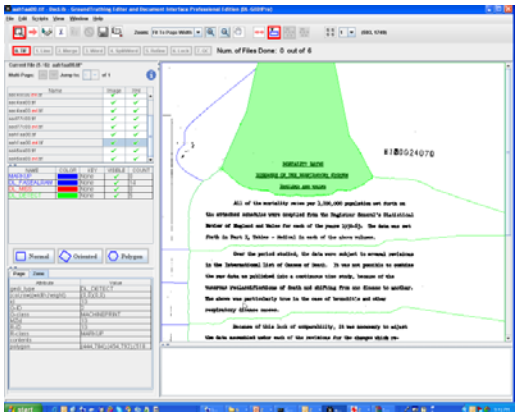
GEDI Integration and Enhancements

- Demo of Version 2.0.2



Agenda

- Review of Goals
- Progress on:
  - Datasets
  - Evaluation Methodology
  - Segmentation Survey and Tools
- Open Discussion of Additional Plans



U.S. Army Research Laboratory

Survey of Page Segmentation and Evaluation Algorithms

Mudit Agrawal

David Doermann

Page Segmentation Algorithms

- Geometric
  - Dividing document into homogenous zones
- Layout
  - Providing Zone content labeling
  - Assigning logical relations based on location

Focus

- Identify the primary segmentation Algorithms
  - Quick overview of each
- Identify likely candidates for Segmentation of Anfal Data

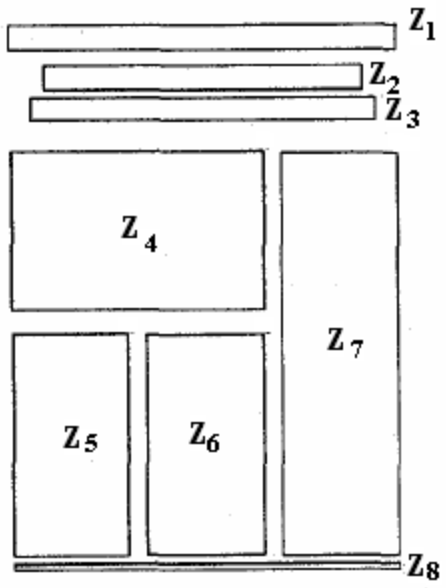
- NOTE:
  - Anfal type line finding is supported by MadCat….

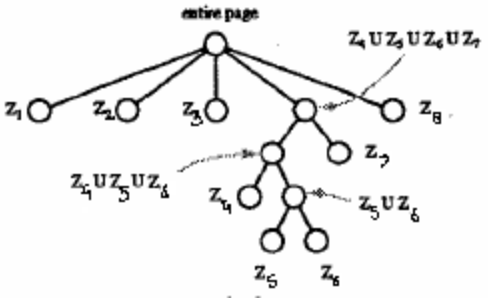
Geometric Page Segmentation

- X-Y cuts
- Smearing
- Whitespace Analysis
- Constrained Text-Line Detection
- Docstrum
- Voronoi based

Recursive X-Y cuts

- At each step, the pixel projection profiles are calculated in both horizontal and vertical directions
- Zone division is performed at most prominent valley in either projection profile
- Process is repeated recursively until no sufficient wide valleys are left in either profile





Smearing

(a) Original Image

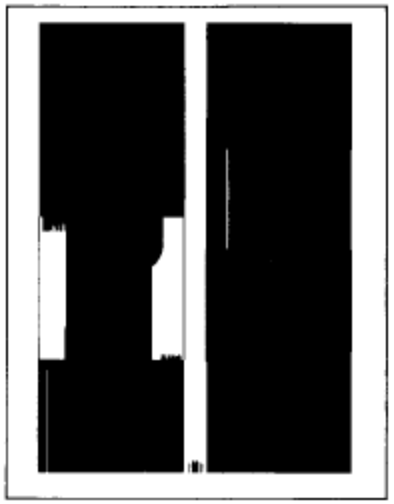
(b) (c) Smearing in Horizontal & Vertical Directions with different Thresholds

(d) Combining using AND operation

(e) Text regions

(a)

(b)



(c)

(d)

(e)

U.S. Army Research Laboratory

Whitespace Analysis

- Find a set of maximal white rectangles (covers)
- Covers are sorted by

$$K(c) = \sqrt{\text{area}(c) * W(|log_2\,(\text{height}(c)/\text{width}(c))|)}$$

- Weighing function assigns higher weights to tall and long rectangles
- Covers are combined one by one (as per their weights)
- A segmentation is the uncovered area left by the union of the covers combined so far

UNIVERSITY OF 18 56 MARYLAND

U.S. Army Research Laboratory

Constrained Text-Line Detection

- Only needs to find a list of obstacles that lines of text do not cross
- Obstacles = gutters, e.g. figures or thin vertical lines
- Tall whitespace rectangles, column separators are candidates for gutters
- Using a robust least square method, contribution of each character to the overall match score of a text-line is penalized by the square of the distance of the alignment point from the base line

Docstrum

- Connected components are separated into two groups (using size ratio factor f_d)
  - Dominant characters
  - Characters in titles and section headings
- For each connected component, K nearest neighbors are found
- Text-lines are computed using transitive closure on within-line nearest neighbor pairings (threshold f_t)
- Text-lines are merged using parallel and perpendicular distance thresholds to form blocks

Voronoi Based Segmentation

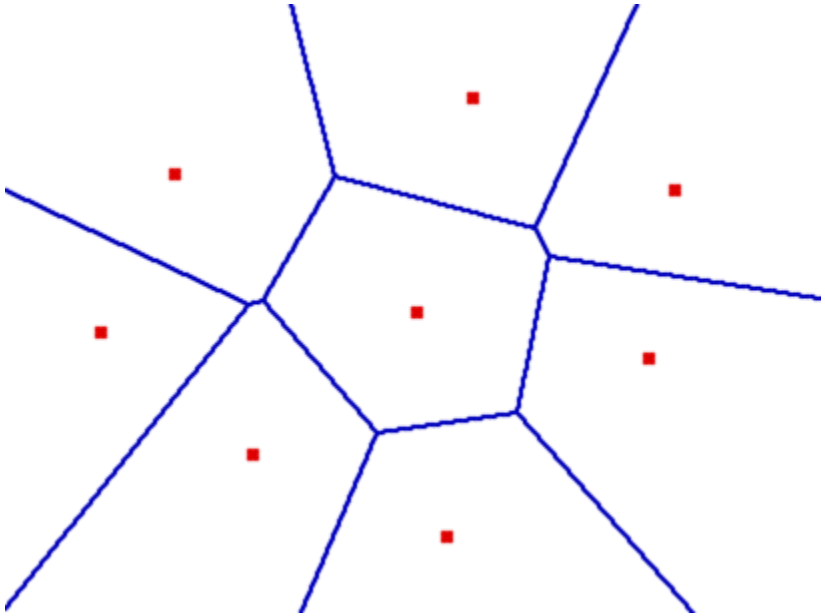
- Based on iterative removal of partitions
- Can be trained
- Can be extended to consider context
- Can be made robust to noise

Options for Arabic?

• X-Y cuts	Layout too Complex
• Smearing	Layout too Complex
• Whitespace Analysis	Noisy
• Constrained Text-Line	More Types of Zones
• Docstrum	Zone Overlap
• Voronoi based	Maybe

Step 1

Point Voronoi Diagram



Voronoi Region of point p_i

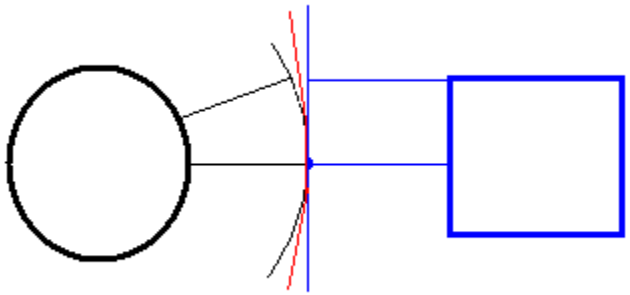
$$V(p_i) = \{p \mid d(p, p_i) \leq d(p, p_j), \forall j \neq i\}$$

UNIVERSITY OF 18 56 MARYLAND

U.S. Army Research Laboratory

Step 2

Area Voronoi Diagram



Voronoi Region of area g_i

$$V(g_i) = \{p \mid d(p, g_i) \leq d(p, g_j), \forall j \neq i\}$$

where

$$d(p, g_i) = \min_{q \in g_i} d(p, q)$$

- Area Voronoi approximation using Point Voronoi diagram:

  - $P_i = \{p_{i1}, \ldots p_{im}\}$ be a set of points lying on the boundary of a figure g_i
  - Generate point voronoi from generators $P = P_1 \cup P_2 \ldots \cup P_n$
  - For all *i,j,k* delete voronoi edges from points of same figure, i.e. p_{ij} and p_{ik}

U.S. Army Research Laboratory

Procedure

- Labeling
- Border Following
- Sampling rate [sr]
- Create area voronoi diagram using sampled points
- Select appropriate Voronoi edges
  - Min distance
  - Area ratio

Features for selection

- Min Distance

$$d(E) = \min_{1 \le i \le m} d(p_i, q_i)$$

where

p_i & q_i are pair of points constituting ith edge between CCs

- Area Ratio

$$a_r(E) = \frac{\text{max of areas of 2 CCs}}{\text{min of areas of 2 CCs}}$$

- Delete an edge if
  - $d(E)/T_{d1} < 1$
  - $d(E)/T_{d2} + a_r(E)/T_a < 1$

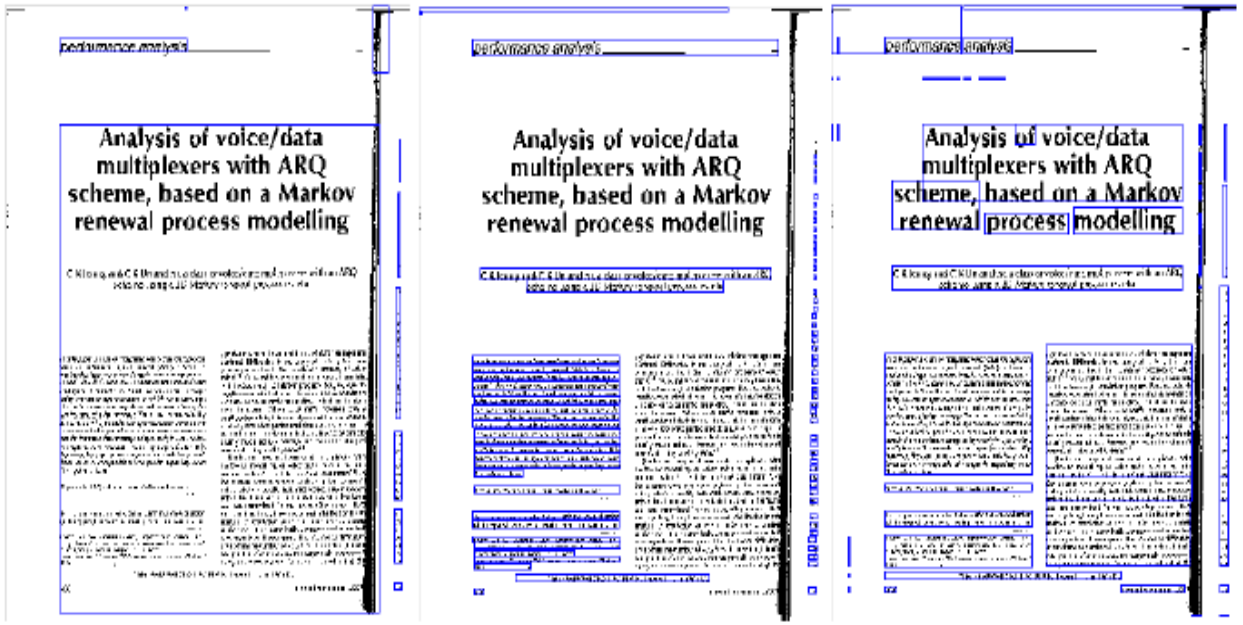
  where $T_{d1} < T_{d2}$

UNIVERSITY OF 18 56 MARYLAND

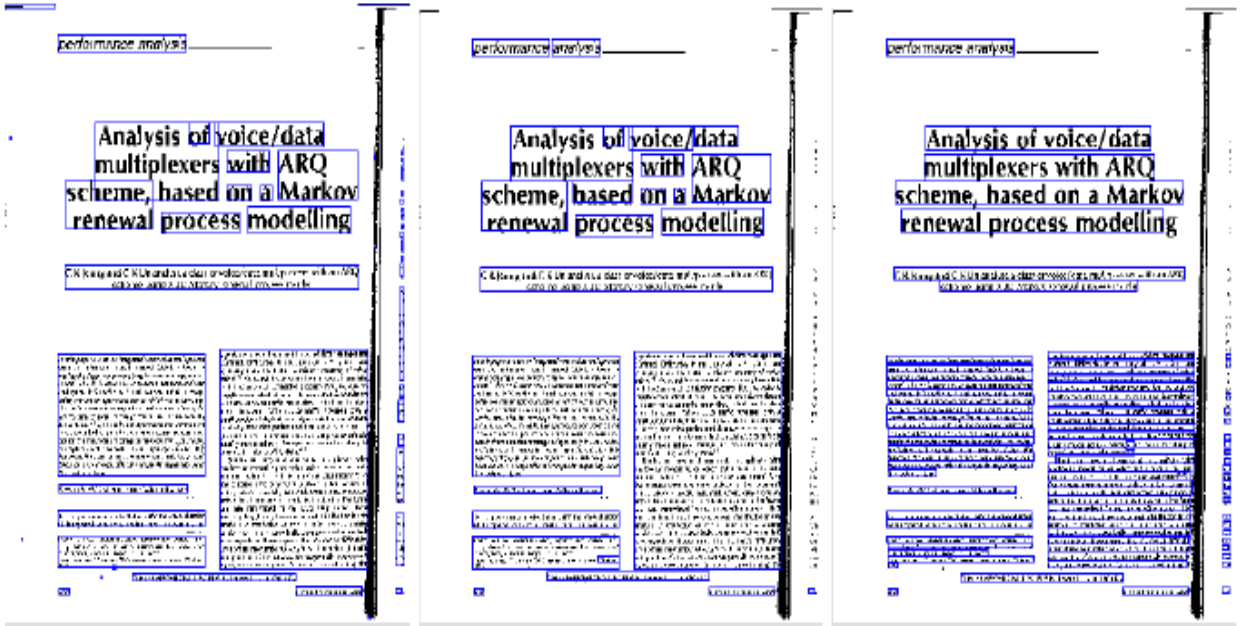
U.S. Army Research Laboratory

Parameters

Parameter	Description	Sensitive (Y/N)?
sr	Sampling rate	Y
nm	Size Th on noise CC	Y
Ch	CC height Th	N
Cw	CC width Th	N
Cr	CC aspect ratio Th	N
Az	Min area Th of a zone	N
Br	Max aspect ratio Th	N
sw	Smoothing window	N
Td1	Inter char Th1	Y
Td2	Inter char Th2	Y
Ta	Area ratio Th	Y



(a) X-Y cut (b) Smearing (c) Whitespace



(d) Docstrum (e) Voronoi (f) Text-line

UNIVERSITY OF MARYLAND 18 56

U.S. Army Research Laboratory

Error Measurements & Metric Definitions

- Ground-truth data had only text-line blocks

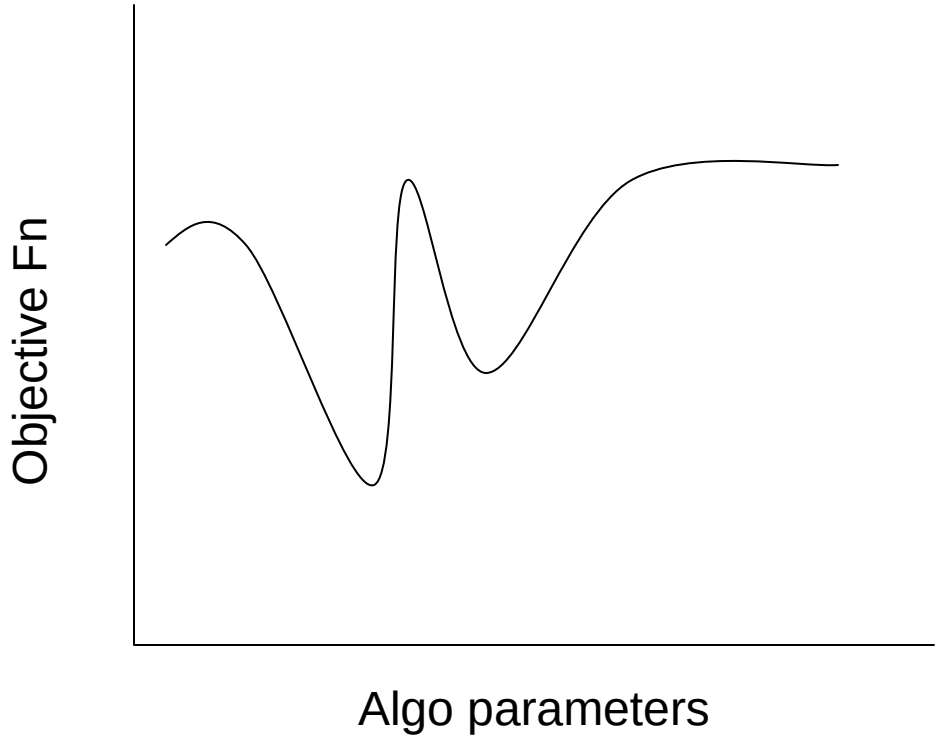
- Three types of textline based error metrics
  - Ground-truth textlines that are *missed*
  - GT textlines whose bounding box is *split*
  - GT textlines that are horizontally *merged*

$$\rho(I, G, R) = \frac{\#\mathcal{L} - \#\{C_L \cup S_L \cup M_L\}}{\#\mathcal{L}}.$$

where

C_L missed
S_L split
M_L merged

Training of Page Segmentation Algorithms



UNIVERSITY OF MARYLAND 18 56

U.S. Army Research Laboratory

Objective Function

Minimizing the objective function:

$$f(\mathbf{p}^A; \mathcal{T}, A, \rho) = \frac{1}{\#\mathcal{T}} \left[\sum_{(I,G)\in\mathcal{T}} 1 - \rho(G, Seg_A(I, \mathbf{p}^A)) \right]$$

where

p^A is parameter vector for A

A is segmentation algorithm

τ is a training dataset

ρ is performance metric as textline accuracy

I is document image

G is ground - truth

UNIVERSITY OF MARYLAND 18 56

U.S. Army Research Laboratory

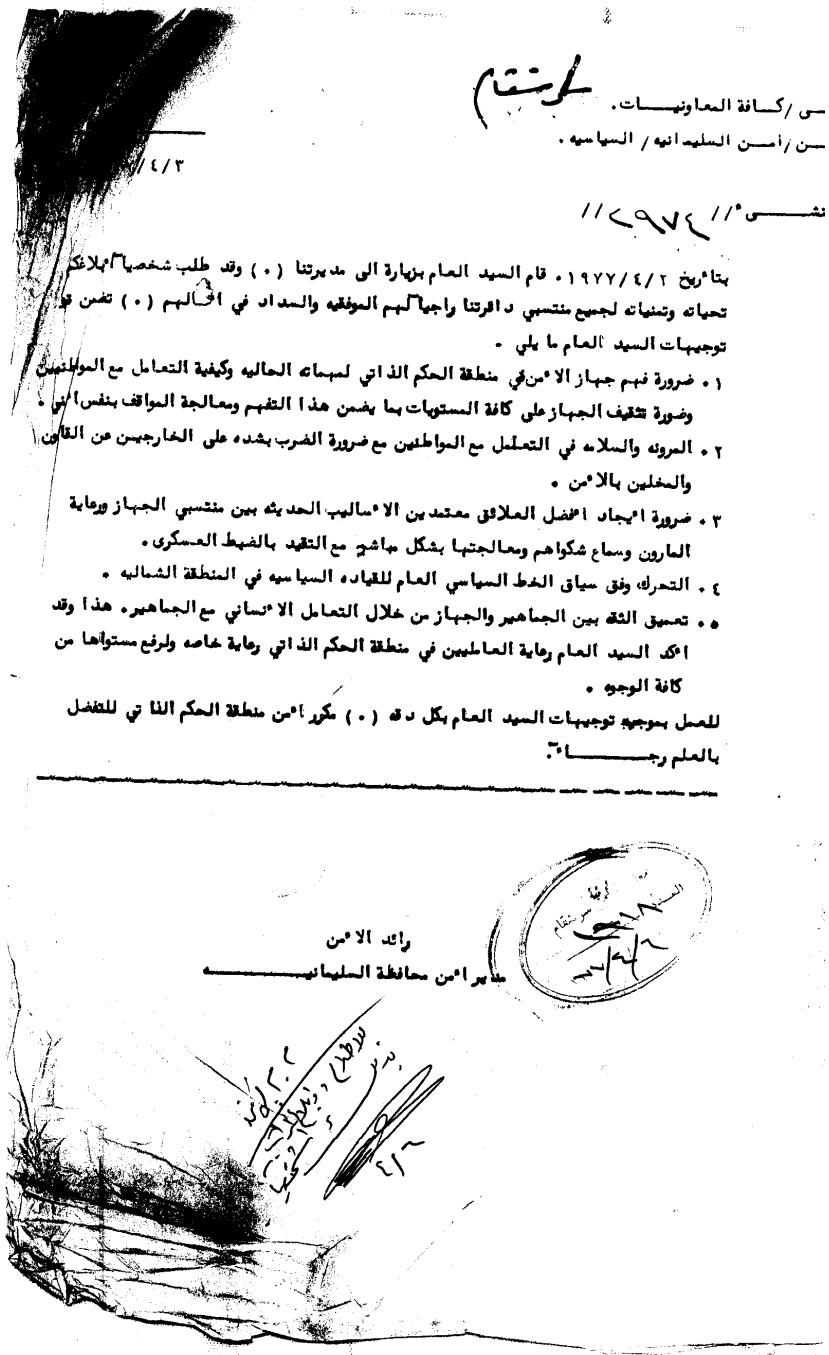
Percentage of different types of errors made by each algorithm

Algorithm	Default parameters			Optimized parameters		
	Split	Merge	Missed	Split	Merge	Missed
Dummy	0.0	65.5	0.0	0.0	65.5	0.0
X-Y cut	5.6	7.8	0.4	5.6	7.8	0.4
Smearing	3.8	1.0	5.7	3.8	1.0	5.7
Whitespace	6.6	1.3	0.0	5.0	2.6	0.0
Text-line	5.1	1.3	0.2	5.1	1.3	0.2
Docstrum	4.5	9.0	0.0	2.5	3.6	0.01
Voronoi	4.9	0.8	0.02	2.9	1.3	0.02

Challenges in Handwriting Documents

- Curvilinear text lines and small or missing linear inter-line gaps
- Stray marks which make rectangular white space analysis difficult
- Local skew
- No well-defined baselines
- Regions not rectangular in nature, hence bounding box may not be the best representation

ANFAL DATA



سي /كافة المعاونيات.
سن /أمن السليمانيه / السياسيه .

نشي/ /٧٩٧٤//

٤/٣/

بتاريخ ١٩٧٧/٤/٢ . قام السيد العام بزيارة الى مديرتنا (٠) وقد طلب شخصياً ابلاغكم تحياته وتمنياته لجميع منتسبي دائرتنا راجياً لهم الموفقيه والسداد في اعمالهم (٠) تضمن توجيهات السيد العام ما يلي .

١ . ضرورة فهم جهاز الامن في منطقة الحكم الذاتي لمهماته الحاليه وكيفية التعامل مع المواطنين وضرورة تثقيف الجهاز على كافة المستويات بما يضمن هذا التفهم ومعالجة المواقف بنفس الوقت .

٢ . المرونه والسلامه في التعامل مع المواطنين مع ضرورة الضرب بشده على الخارجين عن القانون والمخلين بالامن .

٣ . ضرورة ايجاد افضل العلاقات معتمدين الاساليب الحديثه بين منتسبي الجهاز ورعاية المارون وسماع شكواهم ومعالجتها بشكل مباشر مع التقيد بالضبط العسكري .

٤ . التحرك وفق سياق الخط السياسي العام للقياده السياسيه في المنطقة الشماليه .

٥ . تعميق الثقه بين الجماهير والجهاز من خلال التعامل الانساني مع الجماهير . هذا وقد اكد السيد العام رعاية العاملين في منطقة الحكم الذاتي رعاية خاصه ولرفع مستواهم من كافة الوجوه .

للعمل بموجبه توجيهات السيد العام بكل دقه (٠) مكرر امن منطقة الحكم الذاتي للتفضل بالعلم رجـــــاءً .

رائد الامن
مدير امن محافظة السليمانيـــــــه

امر اداري

بالتحري لثبوت مخالفة ابن ادريس قاسم رشيد فقد [illegible] حركة مقفلة بدون علم ضابط ان [illegible] حيث معافية بقطع خط فارغ لمدة ثلاثة ايام استنادا لاحكام المادة ٦٩ من [illegible] المادة الثامنة من قانون اصول المحاكمات العسكرية .

الى

صورة حسن
معاونية امن السليمانية - كتابيه ٢٦٦ في ٢٤/٧/٧٧ للعلم
صحبة المادة سلام والاحترام
سيتم اجراء اللازم بحقه
امر الفوج
المقدم الفوج

[redacted]

سلمنتقام

لـــى /كـافة المعاونيـــات.

ـــن /امـــن السليمانيه / السياسيه.

/٤/٣

لمنشـــى // ٢٩٧٤ //

بتاريخ ٢/٤/١٩٧٧ . قام السيد العام بزيارة الى مديرتنا (.) وقد طلب شخصياً ابلاغكم تحياته وتمنياته لجميع منتسبي دائرتنا راجياً لهم الموفقيه والسداد في اعمالهم (.) تضمن توجيهات السيد العام ما يلي .

١ . ضرورة فهم جهاز الامن في منطقة الحكم الذاتي لمهماته الحاليه وكيفية التعامل مع المواطنين وضرورة تثقيف الجهاز على كافة المستويات بما يضمن هذا التفهم ومعالجة المواقف بنفس الامن .

٢ . المرونه والسلامه في التعامل مع المواطنين مع ضرورة الضرب بشده على الخارجين عن القانون والمخلين بالامن .

٣ . ضرورة ايجاد افضل العلائق معتمدين الاساليب الحديثه بين منتسبي الجهاز ورعاية المارون وسماع شكواهم ومعالجتها بشكل مباشر مع التقيد بالضبط العسكري .

٤ . التحرك وفق سياق الخط السياسي العام للقياده السياسيه في المنطقة الشماليه .

٥ . تعميق الثقه بين الجماهير والجهاز من خلال التعامل الانساني مع الجماهير . هذا وقد اكد السيد العام رعاية العاملين في منطقة الحكم الذاتي رعاية خاصه ولرفع مستواها من كافة الوجوه .

للعمل بموجبه توجيهات السيد العام بكل دقه (.) مكرر امن منطقة الحكم الذاتي للتفضل بالعلم رجـــــاءً .

رائد الامن

مدير امن محافظة السليمانيـــــــــه

سلمنتقام

لـــى /كـافة المعاونيـــات.

ـــن /امـــن السليمانيه / السياسيه.

/٤/٣

لمنشـــى // ٢٩٧٤ //

بتاريخ ٢/٤/١٩٧٧ . قام السيد العام بزيارة الى مديرتنا (.) وقد طلب شخصياً ابلاغكم تحياته وتمنياته لجميع منتسبي دائرتنا راجياً لهم الموفقيه والسداد في اعمالهم (.) تضمن توجيهات السيد العام ما يلي .

١ . ضرورة فهم جهاز الامن في منطقة الحكم الذاتي لمهماته الحاليه وكيفية التعامل مع المواطنين وضرورة تثقيف الجهاز على كافة المستويات بما يضمن هذا التفهم ومعالجة المواقف بنفس الامن .

٢ . المرونه والسلامه في التعامل مع المواطنين مع ضرورة الضرب بشده على الخارجين عن القانون والمخلين بالامن .

٣ . ضرورة ايجاد افضل العلائق معتمدين الاساليب الحديثه بين منتسبي الجهاز ورعاية المارون وسماع شكواهم ومعالجتها بشكل مباشر مع التقيد بالضبط العسكري .

٤ . التحرك وفق سياق الخط السياسي العام للقياده السياسيه في المنطقة الشماليه .

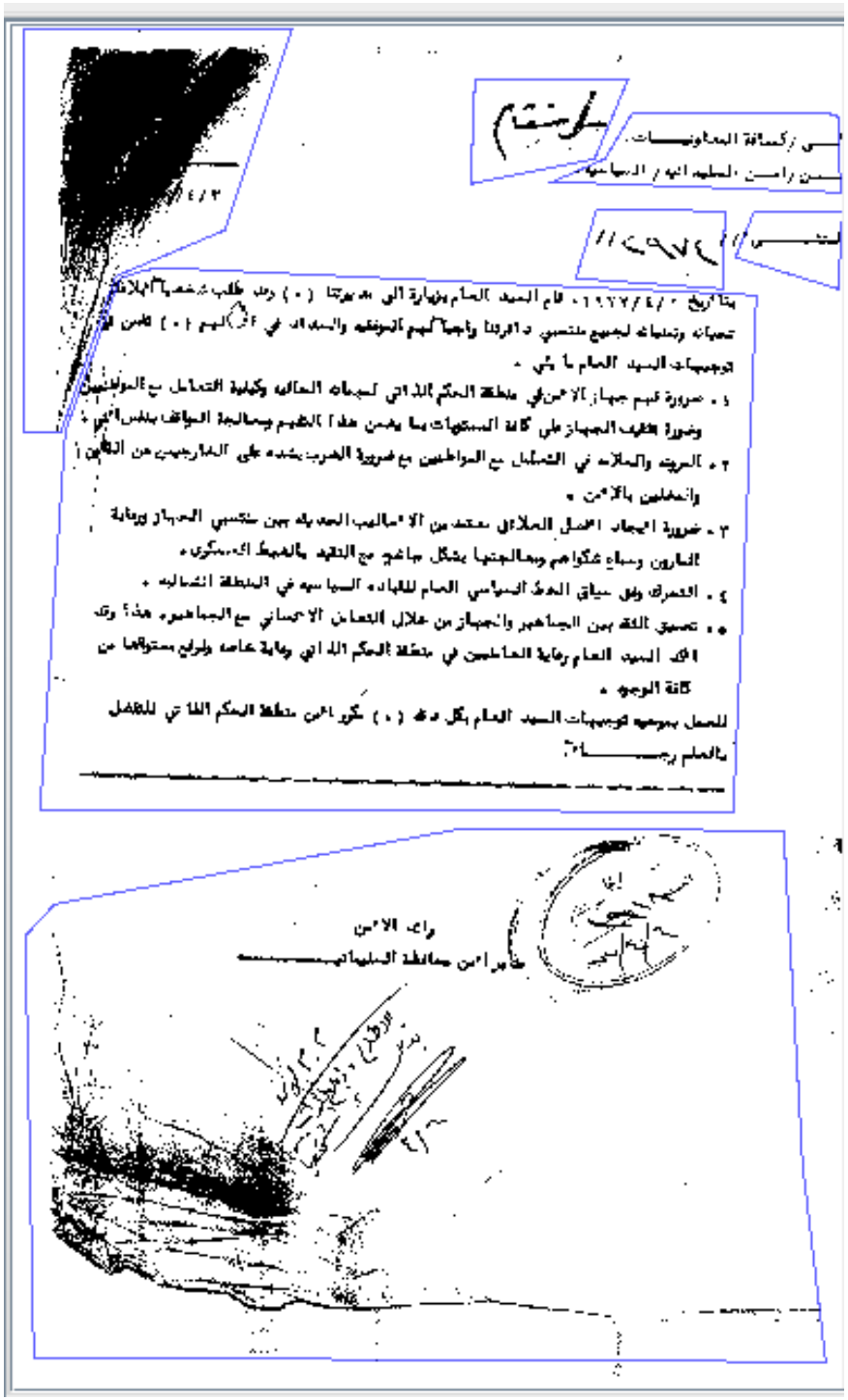
٥ . تعميق الثقه بين الجماهير والجهاز من خلال التعامل الانساني مع الجماهير . هذا وقد اكد السيد العام رعاية العاملين في منطقة الحكم الذاتي رعاية خاصه ولرفع مستواها من كافة الوجوه .

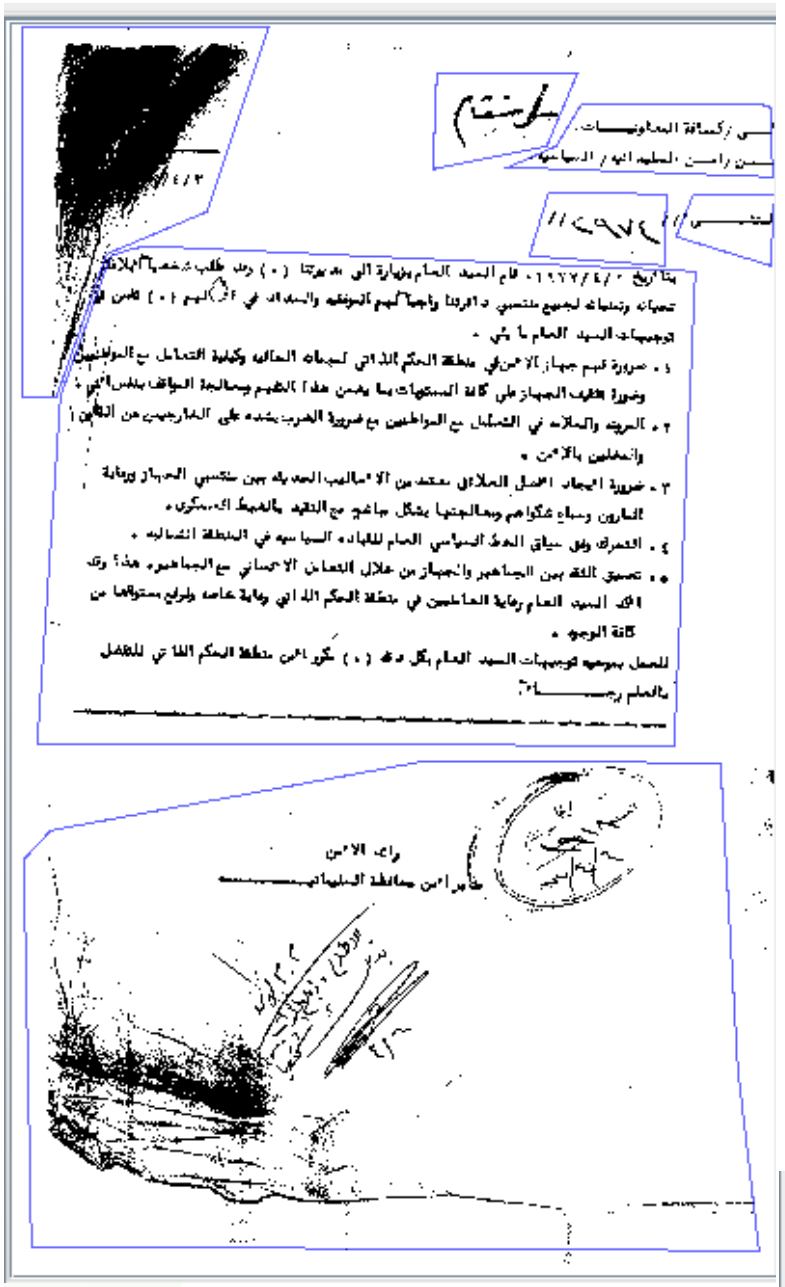
للعمل بموجبه توجيهات السيد العام بكل دقه (.) مكرر امن منطقة الحكم الذاتي للتفضل بالعلم رجـــــاءً .

رائد الامن

مدير امن محافظة السليمانيـــــــــه

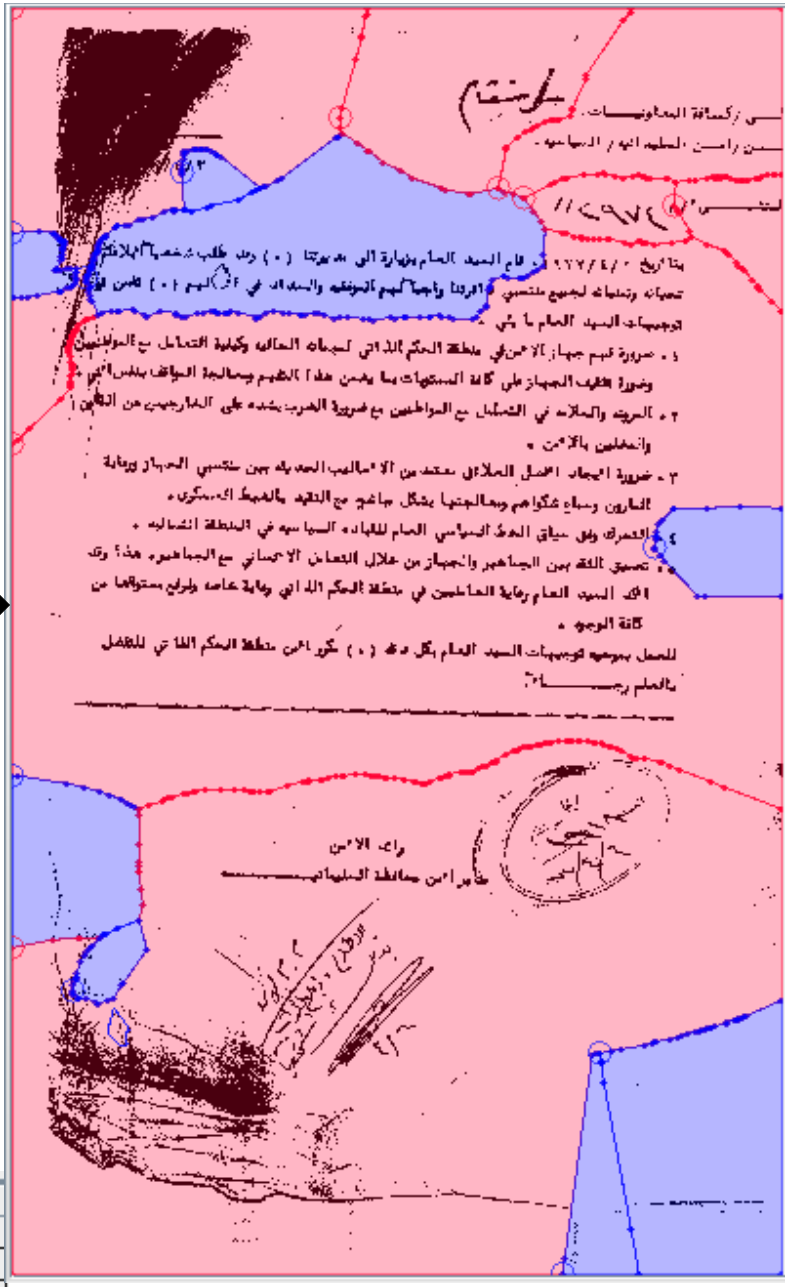
ANFAL DATA





← GT

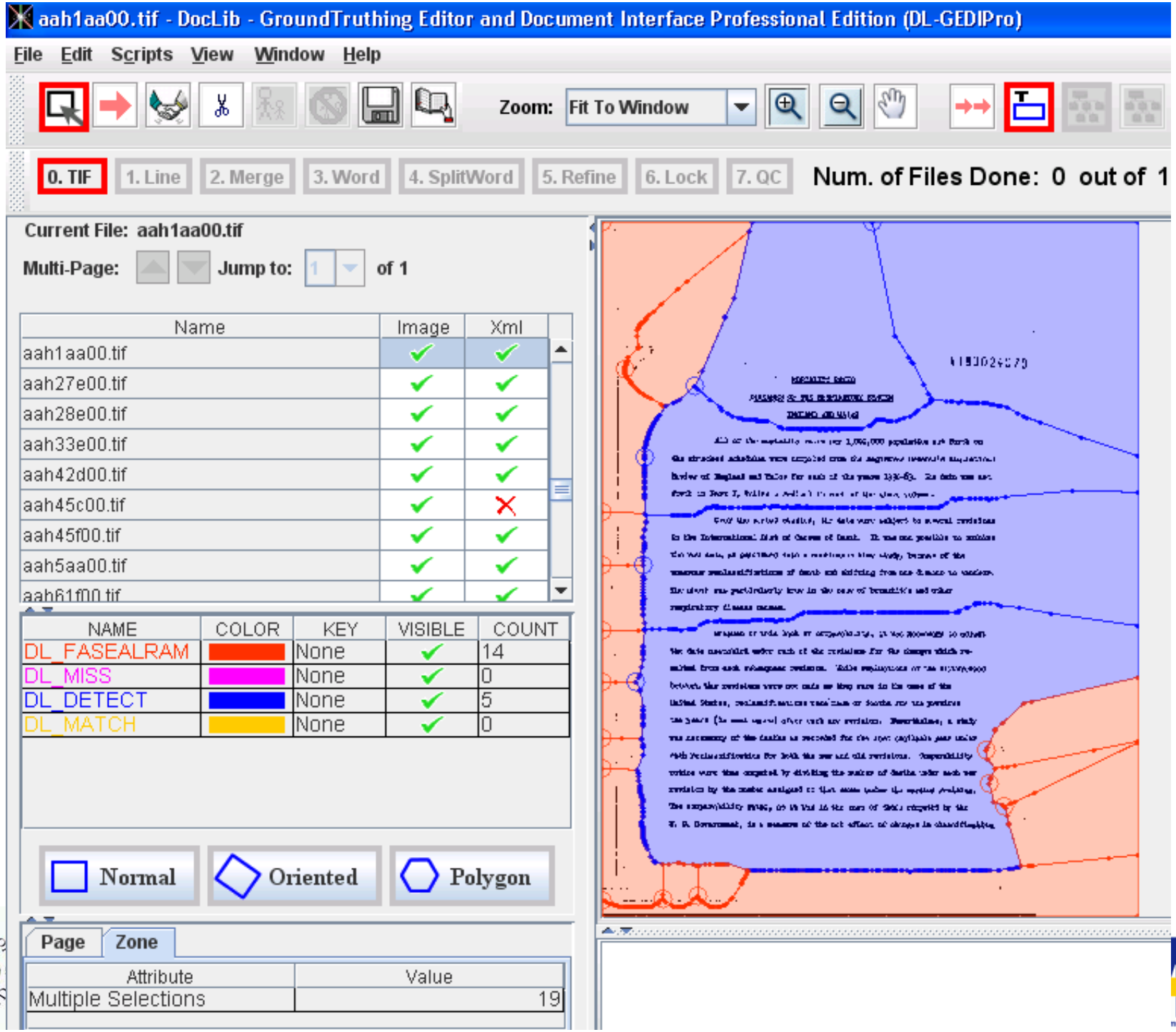
Evaluation →



NAME	COLOR
MATCHED	
DL_FASEALRAM	
FALSEALRAM	
DL_DETECT	
DL_MISS	
MARKUP	
MISSED	

UNIVERSITY OF MARYLAND 18 56

U.S. Army Research Laboratory


aah1aa00.tif - DocLib - GroundTruthing Editor and Document Interface Professional Edition (DL-GEDIPro)
File Edit Scripts View Window Help
Zoom: Fit To Window
0. TIF 1. Line 2. Merge 3. Word 4. SplitWord 5. Refine 6. Lock 7. QC
Num. of Files Done: 0 out of 1
Current File: aah1aa00.tif
Multi-Page: Jump to: 1 of 1
Name Image Xml
aah1aa00.tif
aah27e00.tif
aah28e00.tif
aah33e00.tif
aah42d00.tif
aah45c00.tif
aah45f00.tif
aah5aa00.tif
aah61f00.tif
NAME COLOR KEY VISIBLE COUNT
DL_FASEALRAM None 14
DL_MISS None 0
DL_DETECT None 5
DL_MATCH None 0
Normal
Oriented
Polygon
Page Zone
Attribute Value
Multiple Selections 19

Zone Classification

```
==================
Summary of Results
==================

- Total Number of Sample :  21786
- Overall Accuracy : 95.78%
- Average of Each Class Accuracy : 55.31%

01. Information on Classes
==========================
```

Label	Name of Class	Number of Sample	Accuracy
00	text_sm	20617	97.34%
01	ruling	201	61.69%
02	drawing	299	88.29%
03	table	76	46.05%
04	text_lg	51	64.71%
05	math	301	60.47%
06	halftone	144	83.33%
07	logo	13	0.00%
08	chm_drawing	80	51.25%
09	map	4	0.00%

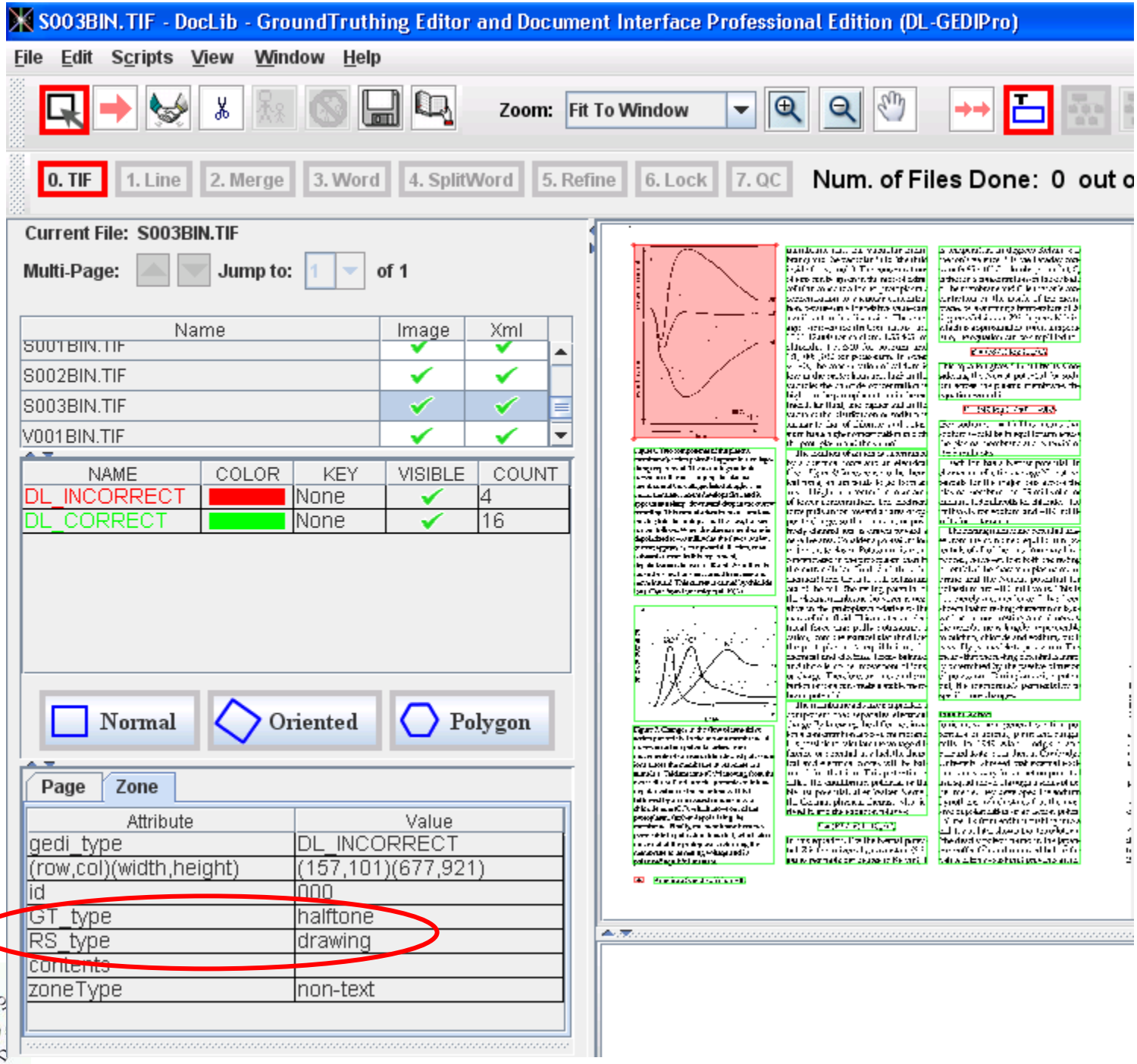
02. Confusion Matrix

Out\GT	00	01	02	03	04
00	20068(97.3%)*	70(34.8%)	11(3.7%)	14(18.4%)	12(23.5%)
01	69(0.3%)	124(61.7%)*	0(0.0%)	1(1.3%)	1(2.0%)
02	93(0.5%)	1(0.5%)	264(88.3%)*	23(30.3%)	4(7.8%)
03	46(0.2%)	0(0.0%)	5(1.7%)	35(46.1%)*	0(0.0%)
04	19(0.1%)	1(0.5%)	0(0.0%)	0(0.0%)	33(64.7%)*
05	284(1.4%)	2(1.0%)	8(2.7%)	2(2.6%)	1(2.0%)
06	38(0.2%)	3(1.5%)	6(2.0%)	0(0.0%)	0(0.0%)
07	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)
08	0(0.0%)	0(0.0%)	5(1.7%)	1(1.3%)	0(0.0%)
09	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)

05	06	07	08	09
106(35.2%)	5(3.5%)	7(53.8%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	1(7.7%)	0(0.0%)	0(0.0%)
9(3.0%)	18(12.5%)	0(0.0%)	9(11.3%)	4(100%)
0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	4(30.8%)	0(0.0%)	0(0.0%)
182(60.5%)*	0(0.0%)	0(0.0%)	30(37.5%)	0(0.0%)
0(0.0%)	120(83.3%)*	0(0.0%)	0(0.0%)	0(0.0%)
0(0.0%)	0(0.0%)	0(0.0%)*	0(0.0%)	0(0.0%)
4(1.3%)	1(0.7%)	1(7.7%)	41(51.2%)*	0(0.0%)
0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)*

03. Precision and Recall

Class\Eval	precision	recall	detected	correct	total
00	98.89%	97.34%	20293	20068	20617
01	63.27%	61.69%	196	124	201
02	62.12%	88.29%	425	264	299
03	40.70%	46.05%	86	35	76
04	57.89%	64.71%	57	33	51
05	35.76%	60.47%	509	182	301
06	71.86%	83.33%	167	120	144
07	0.00%	0.00%	0	0	13
08	77.36%	51.25%	53	41	80
09	0.00%	0.00%	0	0	4


S003BIN.TIF - DocLib - GroundTruthing Editor and Document Interface Professional Edition (DL-GEDIPro)
File Edit Scripts View Window Help
Zoom: Fit To Window
0. TIF 1. Line 2. Merge 3. Word 4. SplitWord 5. Refine 6. Lock 7. QC
Num. of Files Done: 0 out o
Current File: S003BIN.TIF
Multi-Page: Jump to: 1 of 1
Name Image Xml
S002BIN.TIF
S003BIN.TIF
V001BIN.TIF
NAME COLOR KEY VISIBLE COUNT
DL_INCORRECT None 4
DL_CORRECT None 16
Normal
Oriented
Polygon
Page Zone
Attribute Value
gedi_type DL_INCORRECT
(row,col)(width,height) (157,101)(677,921)
id 000
GT_type halftone
RS_type drawing
contents
zoneType non-text

Remaining Tasks

- Evaluation of Existing Data
- Sponsor testing of software
- Integration of OCR evaluation
- Feedback from MADCAT Participants