BOBCAT UMD Progress Report March 28, 2008

Task: <u>Metrics</u> Researchers: Wontaek Seo, David Doermann

Completed Work:

Algorithms and a frame work has been designed and implemented for document zone classification evaluation. The program takes to DocLib XML files and compares them on a zone by zone basis, assuming zoneids match, and produces

- A file by file evaluation, showing the zones which are correct and incorrect (See Appendix A)
- A summery of accuracy by zone type (See Appendix A)
- A confusion matrix (See Appendix A)
- A Visual output in XML format showing the correct and incorrect images overlaid on the (See Appendix B)

Status:

We are currently implementing a doclib module for zone segmentation evaluation found in the following ICDAR article.

 Antonacopoulos, A.; Gatos, B.; Karatzas, D., "ICDAR 2003 page segmentation competition," *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, vol., no., pp. 688-692, 3-6 Aug. 2003

We are adding capabilities for zone class results, precision and recall. A short description of the approach, taken from the original article is contained in Appendix C.

Planned Work: Challenges and Issues:

None at this time

Task:Survey of Available data and MetricsResearchers:Kamal Tayal

Status:

A www site has been designed and is operational for the collection of information about datasets, tools and metrics for evaluation. The site will collect contact information, particulars of the datasets and metrics and information about availability and cost.

Planned Work:

February 2008 – testing and finishing implementation

March 2008 – open data collection and targeted advertising

Challenges and Issues:

The timeline looks like the delivery will slip a little from the previously specified February 15^{th} .

Task:Ground Truth DataResearchers:David Doermann

Status:

A specification has been designed and a preliminary version will be available at the end of March.

The group intends to take the 200 plus document images from the ANFAL collection that have been transcribed and have additional metadata and do line and zone level ground truth.

We will approach the sponsors about using the MADCAT specifications so the projects can both use the data.

Planned Work: Ongoing Challenges and Issues: None at this time

Task: GEDI EnhancementsResearchers: Elena Zotkina and David Doermann

Previous Work:

The GEDI tool has been modified to provide reading order capabilities. It has also added some automated "box shrinking" capabilities that will be used for zone level ground truthing

Status:

Various bug Fixes Recent enhancements include File Read/Write Warnings Reading Order Capabilites Network listener added to control GEDI from outside the program Linux Support Multiple File Type Support Most recent version Delivered to ARL in March

Planned Work:

Version 2.1 tested and delivered April 15th.

Appendix A: Evaluation Output

Zone classification Evaluation Result Generated on Feb 14 00:20:42 2008

Result of Individual File

O : Matched, X : Mis-matched

A001BIN.TIF

(0) ZoneID : 000, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 001, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 002, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 003, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 004, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 005, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 006, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 007, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 008, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 009, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 009, Ground-Truth : text_sm, Output : text_sm

A002BIN.TIF

(0) ZoneID : 000, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 001, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 002, Ground-Truth : text_sm, Output : text_sm (0) ZoneID : 003, Ground-Truth : table, Output : table [OVERALL] 4/4, 100.00%

<<SKIPPED SOME CONTENT>>

Summary of Results

- Total Number of Sample : 24531

- Overall Accuracy : 94.73%

- Average of Each Class Accuracy : 66.12%

01. Information on Classes

Label	Name of Class	Number of Sample	Accuracy
00	text_sm	21572	95.89%
01	table	204	49.02%
02	math	837	91.40%
03	text lg	127	69.29%
04	halftone	384	96.61%
05	drawing	774	89.53%
06	chm_drawing	168	65.48%
07	logo	16	6.25%
08	ruling	435	97.70%
09	map	14	0.00%

02. Confusion Matrix

Out\GT	00	01	02	03	04	05	06	07	08	09
00	20686(95.9%)*	10(4.9%)	48(5.7%)	16(12.6%)	0(0.0%)	10(1.3%)	6(3.6%)	2(12.5%)	0(0.0%)	2(14.3%)
01	28(0.1%)	100(49.0%)*	0(0.0%)	0(0.0%)	2(0.5%)	8(1.0%)	0(0.0%)	0(0.0%)	1(0.2%)	0(0.0%)
02	411(1.9%)	11(5.4%)	765(91.4%)*	17(13.4%)	0(0.0%)	18(2.3%)	36(21.4%)	2(12.5%)	1(0.2%)	0(0.0%)
03	15(0.1%)	0(0.0%)	0(0.0%)	88(69.3%)*	1(0.3%)	11(1.4%)	0(0.0%)	5(31.3%)	0(0.0%)	0(0.0%)
04	403(1.9%)	2(1.0%)	0(0.0%)	0(0.0%)	371(96.6%)*	27(3.5%)	0(0.0%)	5(31.3%)	8(1.8%)	3(21.4%)
05	21(0.1%)	81(39.7%)	24(2.9%)	3(2.4%)	10(2.6%)	693(89.5%)*	16(9.5%)	0(0.0%)	0(0.0%)	9(64.3%)
06	3(0.0%)	0(0.0%)	0(0.0%)	3(2.4%)	0(0.0%)	7(0.9%)	110(65.5%)*	1(6.3%)	0(0.0%)	0(0.0%)
07	0(0.0응)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	1(6.3%)*	0(0.0%)	0(0.0%)
08	5 (0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	425(97.7%)*	0(0.0%)
09	0(0.0%)	0(0.0%)	0(0.0%)	0(0.0%)	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	99.55%	95.89%	20780	20686	21572
01	71.94%	49.02%	139	100	204
02	60.67%	91.40%	1261	765	837
03	73.33%	69.29%	120	88	127
04	45.30%	96.61%	819	371	384
05	80.86%	89.53%	857	693	774
06	88.71%	65.48%	124	110	168
07	100.00%	6.25%	1	1	16
08	98.84%	97.70%	430	425	435
09	0.00%	0.00%	0	0	1

Appendix B: Snapshot of results visualization

Note Small Text was incorrectly recognized and classified as a table (Shown in Red)

A003BIN.tif - DocLib - GroundTruthing Editor and Docum	nt Interface Professional Edition (DL-GEDIPro)	
le Edit Scripts View Window Help	▼ @ Q [®] % % % % % % % % %	
urrent File: A003BIN.xml		
	100 COMPUTATIONAL INTELLIGENCE	
Name Image Xml	the second state and the second state of the s	MANN
1001BIN	by humans must be performed at the partial-participant of the performance of the performa	ci: McDo
002BIN 🗸 🗸	have desirable time and space characteristics. We hope that this work is a step toward	ef
.003BIN 🗸 🗸	understanding how people generate language as rapidly as they do and building practical	N
004BIN 🗸 🖌 🗐	language-generation systems that achieve similarly impressive speeds.	McKr
005BIN 🗸 🗸		MARRI
206BIN 🖌 🖌		5.
TOOBIN 🗸 🗸	ACKNOW] EDGMENTS	PATTE
T01BIN 🗸 🖌	mining the second is used by a small from the Center for Compline Science at	PATT
	This work was supported in part by a grant from the Center for Cognitive Betenet in	Ľ
NAME COLOR KEY VISIBLE COUNT	The Onio state conversity.	Roget
L CORRECT None 1		DACKS
	REFERENCES	SCHA
		N
	ALLEN, J. 1987. Natural language understanding. Benjamin/Cummings, Monto Park, CA.	SCHER
	AUSTIN, J. L. 1965. How to do things with words. Oxford University Press, Oxford.	SCHEL
	BOBROW, D. G., K. M. KAPLAN, M. KAY, D. A. NORMAN, R. THORSON, and T. PHOOMSON, The Company of the second s	SHAR
	BRADY, M. AND R. BERWICK, Editory, 1983. Computational models of discourse, M.I.T. Press,	E F
	Cambridge, MA.	SEAR.
	CHANDRASEKARAN, B. 1987. Towards a functional architecture for intelligence based on generic	WINC
Normal Oriented O Polygon	intelligence, op. 1183-1192.	
I Norman Vorienteu	1991 Models versus rules, deep versus compiled, content versus form. IEEE expert, April:	
	75-79. Device the second secon	
Page Zone	diagnostic problem-solving. In Developments in expert systems. Edited by M. J. Coombs.	
Attribute Value	Academic Press, New York, pp. 23-34.	
row col)(width beight) (402 1129)(1666 1691)	COHEN, P. and C. R. PERRAULT, 1986. Elements of a plan-based theory of speech acits. In Readings	
1 006	in natural language processing. Edited by B. Groszy, et al. Morgan Kaumaa, Los Rivos, Ch.	
GT_type text_sm	pp. 425–450. FIKES, R. E., P. E. HART, and N. J. NILSSON. 1972. Learning an executing generalized robot plans.	-
RS_type table	Artificial intelligence 3(4): 251-288.	1.
ontents	GAZDAR, G., E. KLEIN, G. PULLUM, and I. SAG. 1985. Generalized phrase structure grammar.	1000
	Harvard University Press, Cambridge, MA.	
	eastern states conference on linguistics. Department of Linguistics, Columbus, Ohio.	-
	————————————————————————————————————	
	annual meeting of the Chicago Linguistics Society. Chicago Linguistics Club, Chicago, pp. 30-	
	64. GEIS, M. L. Forthcoming, Speech acts and social actions, (book ms.)	
	GRISHMAN, R. and R. KITTREDGE. Editors. 1986. Analyzing language in restricted domains. Erlbaum.	
	Hillsdale, NJ.	
elected Zone Info	HALLIDAY, M. A. K. 1978. Language as social semiolic. Edward Ardold, London. HALLIDAY, M. A. K. and R. HASAN, 1989. Language, context, and text, Oxford University Press,	
02,1129)(2068,2820)	New York.	
INCORRECT	LAIRD, J. E., P. S. ROSENBLOOM, A. NEWELL, 1986. Chunking in soar: the anatomy of a general	

Appendix C: Description of Algorithm taken from Original published Paper:

The performance evaluation method used is based on counting the number of matches between the entities detected by the algorithm and the entities in the ground truth [5-7]. We use a global MatchScore table for all entities whose values are calculated according to the intersection of the ON pixel sets of the result and the ground truth (a similar technique is used at [8]).

Let I be the set of all image points, Gj the set of all points inside the j ground truth region, Ri the set of all points inside the i result region, gj the entity of j ground truth, ri the entity of i result, T(s) a function that counts the elements of set s. Table MatchScore(i,j) represents the matching results of the j ground truth region and the i result region. Based on a pixel based approach of [5], and using a global MatchScore table for all entities, we can define the metric

If Ni is the count of ground-truth elements belonging to entity i, Mi is the count of result elements belonging to entity i, and w1, w2, w3, w4, w5, w6 are pre-determined weights, we can calculate the detection rate and recognition accuracy for i entity can be used to define the accuracy.

A performance metric for detecting each entity can be extracted if we combine the values of the entity's detection rate and recognition accuracy.

A global performance metric for detecting all entities can be extracted if we combine all values of detection rate and recognition accuracy. If I is the total number of entities and Ni is the count of ground-truth elements belonging to entity i, then by using the weighted average for all EDMi values we can define the following Segmentation Metric (SM).

Relevant References:

[5] I. Phillips and A. Chhabra, "Empirical Performance Evaluation of Graphics Recognition Systems," IEEE Transaction of Pattern Analysis and Machine Intelligence, Vol. 21, No. 9, pp. 849-870, September 1999.

[6] A. Chhabra and I. Phillips, "The Second International Graphics Recognition Contest -Raster to Vector Conversion: A Report," in Graphics Recognition:

Algorithms and Systems, Lecture Notes in Computer Science, volume 1389, pp. 390-410, Springer, 1998.

[7] I. Phillips, J. Liang, A. Chhabra and R. Haralick, "A Performance Evaluation Protocol for Graphics Recognition Systems" in Graphics Recognition: Algorithms and Systems, Lecture Notes in Computer Science, volume 1389, pp. 372- 389, Springer, 1998.

[8]B.A. Yanikoglu, and L Vincent, "Pink Panther: a complete environment for ground-truthing and benchmarking document page segmentation", Pattern Recognition, volume 31, number 9, pp. 1191-1204, 1994.