Towards Improved Paper-based Election Technology

Elisa H. Barney Smith
ECE Department
Boise State University
Boise, ID, USA
EBarneySmith@BoiseState.edu

Daniel Lopresti
CSE Department
Lehigh University
Bethlehem, PA, USA
lopresti@cse.lehigh.edu

George Nagy
ECSE Department
Rensselaer Polytechnic Institute
Troy, NY, USA
nagy@ecse.rpi.edu

Ziyan Wu
ECSE Department
Rensselaer Polytechnic Institute
Troy, NY, USA
wuz5@rpi.edu

***Abstract*—Resources are presented for fostering paper-based election technology. They comprise a diverse collection of real and simulated ballot and survey images, and software tools for ballot synthesis, registration, segmentation, and ground-truthing. The grids underlying the designated location of voter marks are extracted from 13,315 degraded ballot images. The actual skew angles of sample ballots, recorded as part of complete ballot descriptions compiled with the interactive ground-truthing tool, are compared with their automatically extracted parameters. The average error is 0.1 degrees. These results provide a baseline for the application of digital image analysis to the scrutiny of electoral ballots.**

***Keywords*- elections, challenged Minnesota ballots, ballot data, document segmentation, ballot readers, mark recognition.**

I. Introduction

We report the availability of recently created resources for election ballot analysis, and the results that we have already obtained. Although paper-based election technology has not been a major focus in the ICDAR community, it is closely related to form and table analysis and, more generally, to document image registration, segmentation, classification, information extraction and indexing.

The United States is almost unique among nations in entrusting the conduct of elections to a state and county based partisan (i.e., political party) system. Leaving to others to debate the merits of this system versus uniform country-wide administration of elections by a sitting government, we note only the recent worldwide resurgence of interest in paper-based election technologies. In the USA renewed interest was undoubtedly prompted by Florida's "butterfly ballots" and its "hanging chads" that eventually resulted in the Supreme Court's decisive involvement in the Year 2000 Presidential Election. Since then, several other countries have been insisting on a verifiable paper audit trail (VPAT) and, more specifically, on machine-and-human readable paper ballots.

The use of paper ballots does not, however, guarantee fair, verifiable, fast and efficient elections. In a draft report on Voluntary Voting Systems Guidelines for 2007 [1], the Security and Transparency Subcommittee for the Technical Guidelines Development Committee of the National Institute of Standards and Technology (NIST) observes that the use of paper to provide independent auditing capabilities in elections is entirely practical, but that there are undeniably open technical issues that can and should be addressed.

Some of the limitations of current ballot-reader technology are due to its gradual evolution from optical mark recognition (OMR) and mark-sense readers with discrete photocells. Although virtually all ballot readers now use CCD or CMOS optical scanners with CCD or CMOS arrays, many ballot scanners still simply mimic OMR. Furthermore, audit and recount techniques have barely changed in the one hundred and fifty years since the introduction of the "Australian Secret Ballot." Comprehensive background material on the technical aspects of the electoral process can be found in [2], [3].

Among questions that could be addressed by Document Image Analysis (DIA) research are the accuracy, repeatability and reliability of vote tallies, and improved - i.e., faster and more objective - verification and validation. At a lower level, these processes require better imaging, page registration, target location, vote-mark detection, and greater use of within-ballot context to discriminate vote marks and noise from illegal violations of voter anonymity. We have conducted research on Paper and Electronic Records for Elections: Cultivating Trust (*PERFECT*) since 2007 with the conviction that any resulting improvements can benefit other DIA applications as well [4], [5].

We describe in Sections II and III our data and software in the hope that other research groups can also make use of the resources that we have developed. In Section IV we report our current results on recent real-life ballots that we believe make a strong case for increased DIA attention. The only research that we have found along this line is the simultaneous interactive ballot verification reported in [6].

II. Resources

Since at first we had no source of marked ballots, we generated synthetic marks generated on a real ballot layout. 21 pages were synthesized with the *BallotGen* toolkit [7]. The simulated ballots contained filled ovals, check marks, X's, and dots that were scaled, rotated and lightened or darkened, each at five parameter settings. The parameters for each mark were specified in a spreadsheet and compiled into *BallotGen*'s command-line input.

Recognition results on five of the above ballots (about 300 marks) are described in [8], [9]. Comparison of the results of four preprocessing and mark detection methods on the synthesized digital images and on the corresponding printed and scanned images showed that the primary source of errors were marks with insufficient contrast ($< 25\%$ of "normal" reflectance) for adequate digitization. In contrast to the experiments presented in Section IV, these methods did not utilize the alignment of the target ovals on the page because they were designed to find the best way to discriminate voter marks from other scribbles that could invalidate a ballot. The fraction of marks detected with the best combination of methods ranged from 87% to 90%, but there were many ($\sim$25%) false alarms. The specifications and the synthesized image files are posted on the *PERFECT* website [10].

We have also obtained several hundred real ballots from the 2006 general election and earlier primaries in New York State. Only a dozen of these have been scanned. These images are also posted on the *PERFECT* website.

BallotTool is a collection of software components (including *BallotGen*) integrated in a graphical user interface with versions that run under the Linux and Microsoft Windows operating systems. It was used to generate ballot-like surveys in a series of human-factors experiments on students. Some of these experiments, conducted in conjunction with researchers from the Muhlenberg Institute of Public Opinion, explored to what extent students' predisposition affects their judgment in interpreting ambiguous marks. 125 of these survey forms, scanned into TIFF files with corresponding Ground-Truth, are posted on the *PERFECT* website. Other ballot-like survey forms were collected in Lehigh University courses that introduced students to technical problems related to the electoral process. A total of 188 page images from eleven sets of these surveys (each containing 30 questions and answers) are posted on *PERFECT*.

Viewed in isolation, voter marks are often ambiguous. The ambiguity can often be resolved by looking at the entire ballot, or even by considering just the shape or position of neighboring marks. The value of Bayesian style-based classification of voter marks was demonstrated experimentally using as training and test data some of the surveys filled out by the Lehigh students [11].

PERFECT also carries some publications that present proposals for a camera-based ballot reader and for two methods (Unbiased Visual Audit Display and Homogenous Class Display) to improve interactive ballot verification and validation.

The largest collection on the *PERFECT* website consists of 13,435 ballot page images of 6737 challenged Minnesota ballots (Fig. 1) from the 2008 US general election. Because the experiments reported below are based on this corpus, the provenance and appearance of the ballots are described in greater detail in Section III.

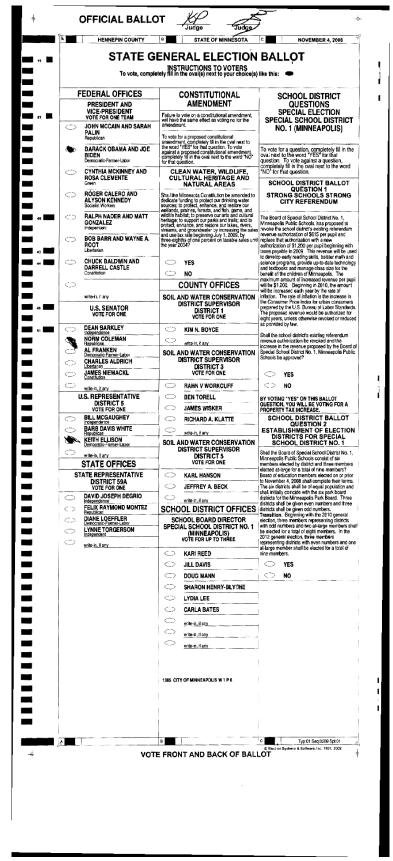


Figure 1. Example of a challenged ballot image (4176 x 1856 pixels)

III. Minnesota Challenged Ballots

During the 2008 General Election in the United States, citizens in the State of Minnesota expressed their preference for U.S. Senator (in addition to U.S. President and dozens of other statewide and local positions). Five senatorial candidates were listed on the ballot. In the initial tally, Republican Norm Coleman received 1,211,590 votes (41.988% of the votes cast) while Democrat Al Franken received 1,211,375 votes (41.981%). Because of the closeness of the race, a mandatory recount was ordered.

In the process of performing recounts, representatives from either candidate could question whether a ballot met the legal requirements set by the State. The reason for the challenge was noted on the (back of the) ballot in the presence of election officials who stamped and initialed the recto. Subsequently the challenged ballots were photocopied, scanned and posted online by Minnesota Public Radio (NPR) (among others) for public comment. The laws that govern the validity of the cast votes, directives for recounts, and the challenge process, are summarized in [12], which also shows many ambiguous examples that divided voting officials and lay voters alike.

To collect all of the ballots from the NPR website, we wrote a simple web crawler that downloaded the files, saving them under their original file names. Another program was then used to extract the images from the PDF and save the recto and verso of each ballot as a separate TIFF file. Examination of the images suggests that most of the ballots were scanned at 300 dpi bitonal, but some of the longer ballots must have been scanned at about 240 dpi (this assessment is based on the assumption that these ballots were printed or copied on 8.5" wide paper), and that they never underwent lossy compression. Hence, they form an ideal dataset for document analysis research.

Challenges occurred in most of Minnesota's 98 counties and 4130 election precincts (including split precincts with several school districts). In a typical precinct, the voter must consider over 75 choices and mark 25 targets for a complete ballot. The layout of the ballots differs from county to county because of differences in the positions up for election, the number of candidates in each race, and the propositions presented for an up-or-down vote.

Although all the ballots are laid out in equally-spaced columns, there are differences in the size and configuration of the headers and instructions in addition to that of the local races. The mandated minimum type size is 6 points for instructions but the candidates' names must be set in at least 10-point type. The races, from Federal Offices to Judicial Offices, and the candidates or propositions within the races, must be listed in a specified order.

Most of the Minnesota ballot templates were provided by two different vendors. One template (Fig. 1) has fiducial marks in the corners (a circle with a cross), uniformly spaced rectangular index bars in the left margin on the recto and in the right margin on the verso, and identical position-coded and numbered identification bars next to the index bars. The other type has smaller index bars along the top and both sides, and position-coded identification bars along the bottom. Many index bars were partially or completely lost in copying or scanning.

Although dedicated ballot scanners may have more reliable paper transports, if the supply at a polling station runs out, ballots may be photocopied. This may add to distortion and degradation introduced by the scanning process.

IV. Experiments on the Minnesota Ballots

The experiments consisted of four different parts carried out separately and at different times: (A) Automated detection of the underlying ballot grid; (B) Interactive entry of all significant information (i.e., ground truth) using *BallotTool*; (C) analysis of the ground truth results; and (D) comparison of ground-truth target oval coordinates with the extracted grid.

A. Extraction of the target grid

Preliminary experiments to locate rulings and the small registration marks ($\oplus$) in the extreme corners showed that they were often missing in the images. The target grid is therefore extracted by locating the alignment of either the solid black horizontal index bars or the empty target ovals. Two different methods are necessary because neither is reliable on all ballots. Excessive skew in photocopying or scanning eliminates most of the index bars because they are located near the edges of the paper. On the other hand, scanner settings that result in very light scans render it impossible to locate the target ovals that even on normal scans have a line thickness of only one or two pixels. In the Challenged Minnesota data, each of these conditions occurs relatively rarely ($< 8\%$), and the combination is, fortunately, very rare ($< 1\%$).

The major processing steps, described in greater detail in [13], are the following

1. Remove rules, text, & noise with adaptive morphological filters
2. Perform connected component (CC) analysis
3. Eliminate CCs with shape or size different from index bars
4. Compute Hough transform to select aligned index bars
5. Select the most populated horizontal and vertical alignments
6. Check spacing, slopes, and alignments against tolerances
7. Switch to oval detection if not enough good index bars
8. Locate CCs with size and aspect ratio of ovals
9. Find near-horizontally and near-vertically aligned ovals
10. Check if within tolerance, otherwise reject
11. Record five grid parameters and plot grid

Some preprocessing is common to both methods. The filter kernels are iteratively adapted to accommodate variations in dpi and in the number of races (1-30) on the page. Furthermore, it is assumed that the approximate size and aspect ratio of the index bars and ovals, and their spacing, are known to within a factor of two because the variation due to different ballot designs and different scanning resolutions falls within this range. All of the ballots have one, two, or three columns of targets. The width of the ballot images ranged from 1136 to 4127 pixels, and their height from 1904 to 5056 pixels. The maximum skew observed was 22°.

The final output consists of five grid parameters $(x_o, y_o, \theta, \Delta x, \Delta y)$ for each page image. The program found enough index bars on 96% of the 13,345 images and rejected fewer than 1% of the pages. Figure 2 displays the range of results. The unoptimized Matlab code ran in 8.3 seconds per page on a 2.83 GHz processor with 32GB RAM. Since we had no ground truth at the time, we sampled 1031 randomly selected pages. All nine of the pages where the program failed were anomalous broken-arrow or absentee ballots. The visually reported accuracy on this 10% sample was in good agreement with the ground-truth based comparison reported below.

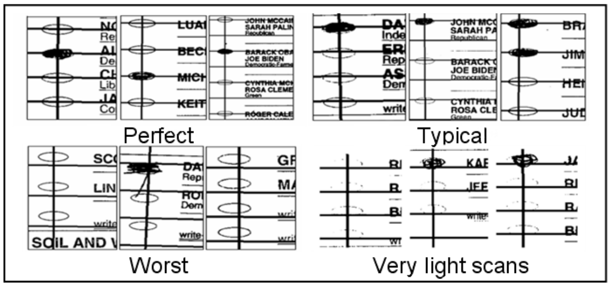


Figure 2. Twelve samples of ballot grids

B. Interactive ballot data entry

A special interface that is part of *BallotTool* is currently being used by teams of students to obtain detailed Ground-Truth for each of these images, Figure 3. So far about 600 ballots have been processed. For each image the operator generates a CSV file indicating the races present on the ballot and the order of the candidates. The skew-angle of the whole page is determined by aligning a ruling line with a line predrawn on the ballot. The location of each marked (oval) target, the presence of overvotes, undervotes, and the position and kind of any other markup added to the original ballot are also identified. The operators also record their opinion about the validity of the vote in each race, and may add freeform notes, comments or observations. Processing of each ballot takes about 10 minutes once the operator is trained. Preparation of each needed CSV file also takes about 10 minutes. New CSV files are needed every 2-10 ballots depending on how many challenged ballots were collected from each voting precinct.

The text files containing the Ground-Truth are posted on *PERFECT*. Part of these keyword-value files have been converted to fixed-format Excel worksheets for greater ease of use (Fig. 4). Some other recorded information cannot, however, be readily accommodated by a spreadsheet.

C. Analysis of ground truth results

There were many reasons these ballots were chosen to be challenged by the political parties. Usually the reason indicated is that the voter did not follow the instructions indicated on the ballot precisely. Of the 620 ballots ground truthed, 444 ballots have handwriting by the officials who reviewed the ballots. 111 of the ballots contain stray marks, 36 have cancelled votes and 116 have handwriting from the voter. On 484 ballots the voter indicated his or her choice with filled ovals as indicated in the instructions. However the voters used partially filled ovals on 173 ballots, x-marks on 30 ballots, check marks on 14 ballots and other marks on 22 ballots to indicate their votes. On 117 ballots the voter used two mark types and on 5 of those ballots three mark types were used.

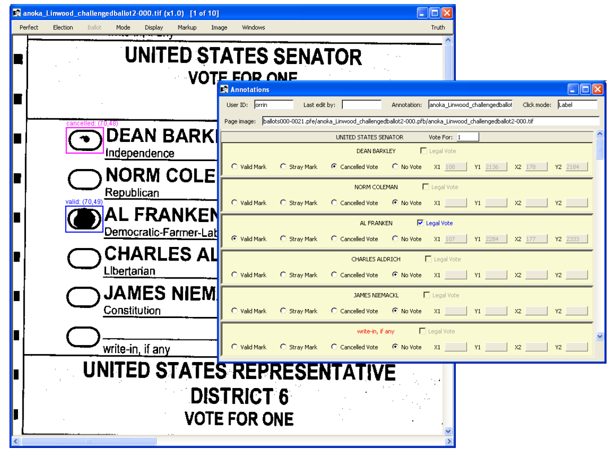
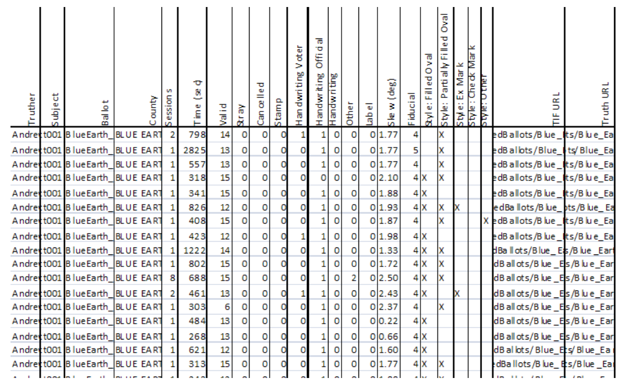


Figure 3. Screen shot of the *BallotTool* interface. A ballot image with location of one valid and one cancelled vote are shown with the screen to indicate what will be drawn on the image.



Truther	Subject	Ballot	County	Sessions	Time (sec)	Valid	Stray	Cancelled	Stamp	Handwriting Voter	Handwriting Official	Handwriting	Other	Label	Skew (deg)	Fiducial	Style: Filled Oval	Style: Partially Filled Oval	Style: Ex Mark	Style: Check Mark	Style: Other	TIF URL	Truth URL
Andre	t001	BlueEarth_	BLUE EART	2	798	14	0	0	0	1	1	0	0	0	1.77	4		X				edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	2825	13	0	0	0	0	1	0	0	0	1.77	5		X				edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	557	13	0	0	0	0	1	0	0	0	1.77	4		X				edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	318	15	0	0	0	0	0	0	0	0	2.10	4	X	X				edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	341	15	0	0	0	0	1	0	0	0	1.88	4	X					edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	826	12	0	0	0	0	1	0	0	0	1.93	4	X	X	X			edBallots/Blue_	ots/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	408	15	0	0	0	0	1	0	0	0	1.87	4		X			X	edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	423	12	0	0	0	1	1	0	0	0	1.98	4	X					edBallots/Blue_	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	1222	14	0	0	0	0	1	0	0	0	1.33	4	X	X				dBallots/Blue_E	s/Blue_Eart
Andre	t001	BlueEarth_	BLUE EART	1	802	15	0	0	0	0	1	0	0	0	1.72	4	X	X				dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	8	688	15	0	0	0	0	1	0	2	0	2.50	4	X	X				dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	2	461	13	0	0	0	1	1	0	0	0	2.43	4	X		X			dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	1	303	6	0	0	0	0	1	0	0	0	2.37	4		X				dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	1	484	13	0	0	0	0	1	0	0	0	0.22	4	X					dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	1	268	13	0	0	0	0	1	0	0	0	0.66	4	X					dBallots/Blue_E	s/Blue_Ear
Andre	t001	BlueEarth_	BLUE EART	1	621	12	0	0	0	0	1	0	0	0	1.60	4	X					dBallots/Blue_E	ts/Blue_Ea
Andre	t001	BlueEarth_	BLUE EART	1	313	15	0	0	0	0	1	0	0	0	1.77	4	X	X				edBallots/Blue_E	ts/Blue_Ear

Figure 4. Recorded Ground Truth for Blue Earth County ballots. In addition to the voter marks, the location and type of extraneous marks, by both the voter and by election officials, are recorded (but not shown here). Some of the longer fields are necessarily truncated.

D. Comparison of extracted grid with ground truth

The groundtruther was asked to indicate the skew of the ballot by using a line tool to select two endpoints of any horizontal rule on the ballot. For 530 of the ballots we have collected this information. When a ballot was ground truthed by multiple people, their skew measurements had a standard deviation less than $0.05°$. The majority of the ballots had less than $0.5°$ absolute skew, Figure 5. The target grid extraction software can also calculate the skew. The automatic results were compared to the manual results. The mean error was $0.12°$ and the standard deviation of the error was $0.24°$.

V. Summary

Obtaining access to hand-marked ballots created by voters for use in real elections has been problematic due to various legal constraints. Fortunately, a recent turn of events created an unprecedented opportunity to address this situation. We

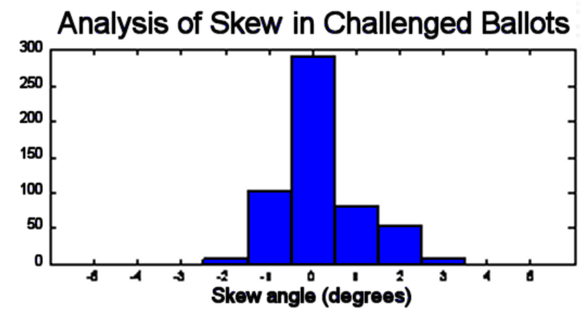


Figure 5. Frequency of skews in ground truthed challenged ballots.

have assembled a large-scale dataset consisting of voter-marked ballot images from the 2008 General Election that was challenged due to the contested Senate Race in the State of Minnesota. So far as we are aware, this is the first such collection ever made openly available and hence is an invaluable resource to those wishing to develop better image processing and pattern recognition methods for reading op-scan ballots.

The ground truthing tool provides information on each ballot that far exceeds what is extracted by commercial ballot counters. In addition to the location and kind of marks, extraneous marks that may lead to invalidating a ballot are recorded. All operator interactions are time stamped in order to provide guidelines for future ground-truthing efforts. We note that reconciling the assigned file names with the correct county and precinct designation was itself a time-consuming task.

We have also provided baseline image processing results that enable other researchers to conduct experiments on improved target location, mark detection, and stray mark characterization with less effort. The automatically extracted grid allows determination of the potential location of the oval targets and vote marks to within 2-3 pixels.

Our main contribution is the assembly of resources and the customization of well-established methods focused specifically on paper ballots. Nevertheless such research, in addition to improving a significant segment of contemporary election technology, may also benefit many other areas (e.g. standardized testing and application forms for various services) that rely on secure mark sensing.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under awards NSF-0716368, NSF-0716393, NSF-0716647, and NSF-0716543. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation. Daniel Lopresti also acknowledges support from a DARPA IPTO grant administered by Raytheon BBN Technologies.

REFERENCES

[1] W. Burr, J. Kelsey, R. Peralta, and J. Wack. Requiring software independence in VVSG 2007: STS recommendations for the TGDC. Technical report, National Institute of Standards and Technology, November 2006.

[2] Marcia Lausen, Design for Democracy, Ballot + Election Design, University of Chicago Press, 2008.

[3] Paul S. Hernson et al., Voting Technology, Brookings Institution Press, Washington, D.C., 2008.

[4] George Nagy and Daniel Lopresti, The role of document image analysis in trustworthy elections, World Scientific Revew Volume (ed. S.K. Parui and B.B. Chaudhuri), in press, 2008-2011.

[5] Daniel Lopresti, George Nagy, Tools for Monitoring, Visualizing, and Refining Collections of Noisy Documents, Procs. Third Workshop on Analytics for Noisy Unstructured Text Data (AND 2009), Barcelona, Spain, 23-24 July 2009.

[6] Arel Cordero, Theron Ji, Alan Tsai, Keaton Mowery, David Wagner, Efficient User-Guided Ballot Image Verification, Procs. USENIX EVT/WOTE, 2010.

[7] Daniel Lopresti, George Nagy, Elisa H. Barney Smith, A Document Analysis System for Supporting Electronic Voting Research, Procs. Document Analysis Systems, Nara, Japan, September 2008.

[8] Elisa H. Barney Smith, Daniel Lopresti, George Nagy, Ballot Mark Detection, Proceedings of International Conference on Pattern Recognition XIX, Tampa, FL, December 2008.

[9] Elisa H. Barney Smith, George Nagy, Daniel Lopresti, Mark Detection from Scanned Ballots, Procs. Document Recognition and Retrieval, San Jose, SPIE, January 2009.

[10] Paper and Electronic Records for Elections: Cultivating Trust (PERFECT), 2010. http://perfect.cse.lehigh.edu/.

[11] Pingping Xiu, Daniel Lopresti, Henry Baird, George Nagy, Elisa H. Barney Smith, Style-Based Mark Recognition, Procs. 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, Spain, 26-29 July 2009.

[12] Daniel Lopresti, George Nagy, Elisa H. Barney Smith, Document Analysis Issues in Reading Optical Scan Ballots, Procs. Document Analysis Systems, Boston, MA, August 2010.

[13] George Nagy, Daniel Lopresti, Elisa H. Barney Smith, Ziyan Wu, Characterizing Challenged Minnesota Ballots, Procs. Document Recognition and Retrieval, San Jose, SPIE, January 2011.