An Analysis of Write-in Marks on Optical Scan Ballots

Theron Ji     Eric Kim     Raji Srikantan     Alan Tsai
Arel Cordero     David Wagner

UC Berkeley
Optical Scan Ballots

- Widely used in today’s elections
- Voters indicate choices by marking voting targets
- Scanner tabulates votes by detecting marks
Write-ins

- Region where write-in candidates are written in by the voter

- Corresponding voting target must be filled for vote to count

- So does this happen?
Motivation

• Lisa Murkowski **wins** the 2010 Alaska Senate election through a write-in campaign

• Donna Frye narrowly **loses** the 2004 San Diego mayoral election because people forgot to mark the write-in voting target
Voter writes in a candidates name, but doesn’t fill in the corresponding voting target – vote is lost

Questions:
- How often does this occur?
- What trends are there when this happens?
- How do you detect this accurately, quickly, and with minimal human effort?
Our Goals

1. Given a large dataset of scanned ballots, develop a system to accurately and efficiently detect write-in marks without using the corresponding voting target.

2. Apply this to a real election and examine the results to see how voters actually use write-in slots on ballots and infer trends or possible sources of error.
Experimental Setup
We were kindly given 248,334 scanned, double-sided ballot images from the 2008 Leon County General Election (thanks to Larry Moore, Ion Sancho, and Clear Ballot Group)

These were in the Premier (Diebold) optical scan format
Assumptions

- We assume we are given blank templates
- We assume ballots have a regular and consistent structure
- (We don’t assume to know write-in locations)
- (We don’t assume scanned image will be perfect)
Image Registration

- Align each ballot to a universal coordinate system
- Necessary for accuracy of further steps
- Robust against folds, skews, and tears in images

[Image of ballot registration process]
Image Registration

- Identify every hashmark along the side using template matching

- OK if some are missing or go undetected
### Image Registration

- **Linear regression along each edge using the hashmarks as points**

- *(Notice the slight leftwards skew in the image as shown by the lines)*
Image Registration

- Correspond every hashmark with the hashmark on the canonical ballot (template)
- Perform an affine transformation
Ballot Grouping

- We group all the ballots of the same style together
- We use the precinct number for this
- Match each style with one of the templates
Finding Write-in Lines

• First we look for the write-in lines

• Notice that they are horizontal lines contained entirely within a contest box

• Use form extraction
Finding Write-in Regions

• Given the write-in lines, we scan upward until whitespace ends

• This gives us a rectangular box that becomes our write-in region
Write-in Detection

- Count the number of black pixels in the write-in region
  - Black Pixels: 8  
  - Black Pixels: 908  
  - Black Pixels: 7203

- Threshold it at a conservative (low) number, and consider anything exceeding the threshold as a mark
Classifying Voting Targets

- Lastly, we classify the voting target for each write-in as filled or unfilled

- Do this through template matching the voting target

- Matched (Unfilled) ×

- Matched (Unfilled) ×

- Not Matched (Filled) ✓
Process Finished…

…Results!
Analysis on the 2008 Leon County General Election
An example task for the participant to do
Common Types of Marks

Actual votes lost

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian

☐ Write-in

☐ Joe Miller

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian

☐ Write-in

☐ Larry Hendricks

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian

☐ Write-in

☐ Sam Bowker

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian

☐ Write-in

STATE REPRESENTATIVE
8TH HOUSE DISTRICT
(Vote for One)

☐ Alan Williams

DEM

☐ Robert Maddox

NPA

☐ Hubert Brown

Write-in
Common Types of Marks

Representative in Congress
2nd Congressional District
(Vote for One)

- ☐ Mark Mulligan
- ☐ Allen Boyd
- ☐ Write-in

Representative in Congress
2nd Congressional District
(Vote for One)

- ☐ Mark Mulligan
- ☐ Allen Boyd
- ☐ Write-in

State Representative
8th House District
(Vote for One)

- ☐ Alan Williams
- ☐ Robert Maddox
- ☐ Write-in

State Representative
8th House District
(Vote for One)

- ☐ Alan Williams
- ☐ Robert Maddox
- ☐ Write-in
Common Types of Marks

**Conflict votes**

- **CITY COMMISSIONER SEAT 1**
  - (Vote for One)
  - Mark Mustian
  - Alvin Akinyemi  
  - Write-in

- **CITY COMMISSIONER SEAT 1**
  - (Vote for One)
  - Mark Mustian
  - Moise Duge
  - Write-in

- **REPRESENTATIVE IN CONGRESS 2ND CONGRESSIONAL DISTRICT**
  - (Vote for One)
  - Mark Mulligan
  - Allen Boyd
  - Robert Ortiz
  - Write-in

- **CITY COMMISSIONER SEAT 1**
  - (Vote for One)
  - Mark Mustian
  - Gregory W. Whitaker
  - Write-in
Common Types of Marks

Non-serious votes 😊

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Bingo TheDog
Write-in

☐ Mark Mustian
☐ Ronald McDonald
Write-in

REPRESENTATIVE IN CONGRESS
2ND CONGRESSIONAL DISTRICT
(Vote for One)

☐ Mark Mulligan
☐ Allen Boyd
Write-in

☐ Mickey Mouse
Write-in

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Jesus Christ
Write-in
Common Types of Marks

Quantifying Votes...

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Write-in

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ None
☐ Write-in

REPRESENTATIVE IN CONGRESS
2ND CONGRESSIONAL DISTRICT
(Vote for One)

☐ Mark Mulligan
☐ Allen Boyd
☐ Neither
☐ Write-in

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Unknown
☐ Write-in
Common Types of Marks

Stray Marks

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Write-in

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Write-in

CITY COMMISSIONER
SEAT 1
(Vote for One)

☐ Mark Mustian
☐ Write-in

STATE REPRESENTATIVE
8TH HOUSE DISTRICT
(Vote for One)

☐ Alan Williams  DEM
☐ Robert Maddox  NPA
☐ Write-in
## Statistics

### Write-in Regions

<table>
<thead>
<tr>
<th>Voting Target</th>
<th>Marked</th>
<th>Unmarked</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filled</td>
<td>834 (0.226%)</td>
<td>78 (0.021%)</td>
<td>911 (0.247%)</td>
</tr>
<tr>
<td>Unfilled</td>
<td>784 (0.213%)</td>
<td>366981 (99.54%)</td>
<td>367766 (99.75%)</td>
</tr>
<tr>
<td>Total</td>
<td>1618 (0.439%)</td>
<td>367059 (99.56%)</td>
<td>368677</td>
</tr>
</tbody>
</table>

\[ \frac{784}{1618} = 48\% \]
Detailed Results

- 1618 write-in votes (834 bubbled, 784 not)
- 453 emphasis votes (3 bubbled, 450 not)
- 17 conflict votes (0 bubbled, 17 not)
- 54 non-serious votes (41 bubbled, 13 not)
- 54 quantifying votes (27 bubbled, 27 not)
- 16 stray marks (0 bubbled, 16 not)

Total Lost votes: **261 (16% of write-in votes)**
Conclusion

• We developed techniques to accurately detect write-in marks from optical scan ballots. We did this with only partial knowledge about the ballot, and minimal human assistance.

• We demonstrated its feasibility on a large, real-life data set from Leon County, and found surprising results – that in fact, up to 16% of write-in votes that could have been counted in the election were lost.
The End

Disclaimer: This was not a real vote