A Document Analysis System for Supporting Electronic Voting Research

Daniel Lopresti

Computer Science & Engineering

Lehigh University

Bethlehem, PA, USA

George Nagy

Electrical, Computer, and Systems Engineering

Rensselaer Polytechnic Institute

Troy, NY, USA

Elisa Barney Smith

Electrical and Computer Engineering

Boise State University

Boise, ID, USA

Overview

- Background: the push toward paper-based voting

- Issues in processing scanned ballots

- Opportunties for document analysis research

- Overview of our ongoing work[*]

- Summary

[*] Ideas and a prototype system, but no experimental results yet.

E-voting in the news

What are the problems?

Recent transition to e-voting in U.S. has been rocky at best:

- Well-publicized "attacks" by computer security researchers who have obtained examples of such systems.

- Votes lost in real elections due to software / hardware failures ...

- ... and due to under-trained workers, bad user interface designs.

> *No matter the vendor, one truth holds:*
> *all computer systems of this complexity have bugs.*

Situation exacerbated by:

- Closed (proprietary) systems, no independent audit trail.

- Result is loss of voter trust, lawsuits, flurry of new legislation.
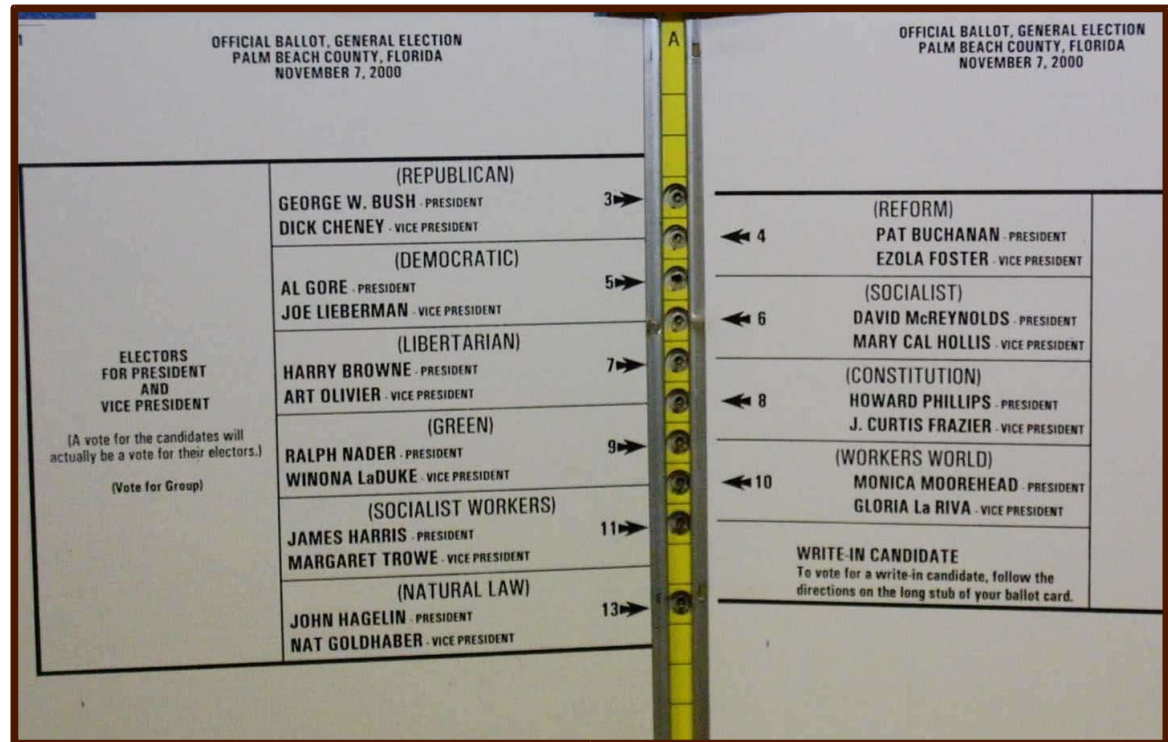
Voting system use in the U.S.
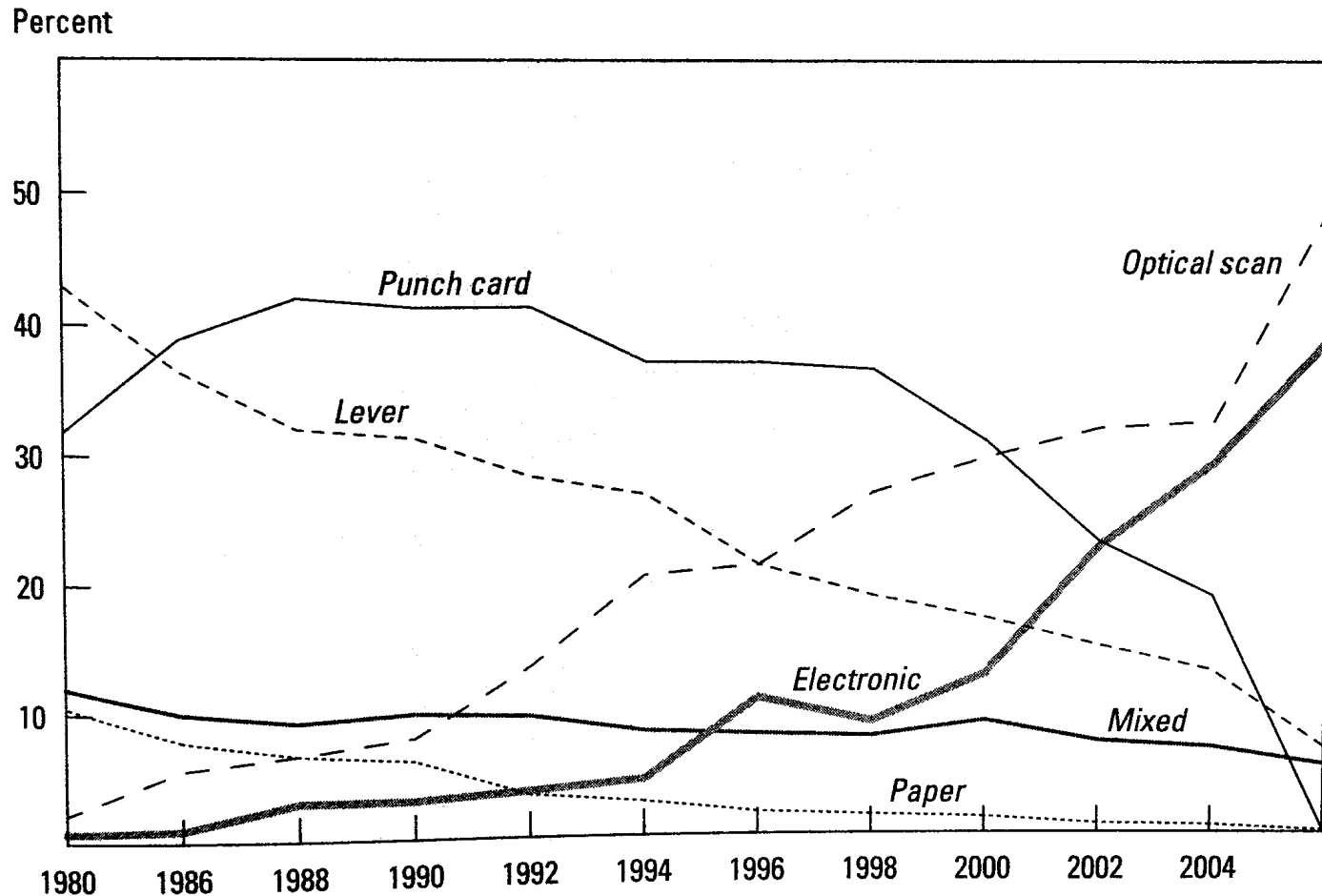
From *Voting Technology: The Not-So-Simple Act of Casting a Ballot*, by Paul S. Herrnson, et al, Brookings Institution Press, 2008.
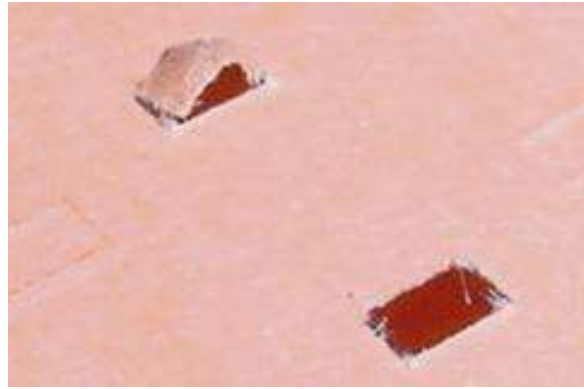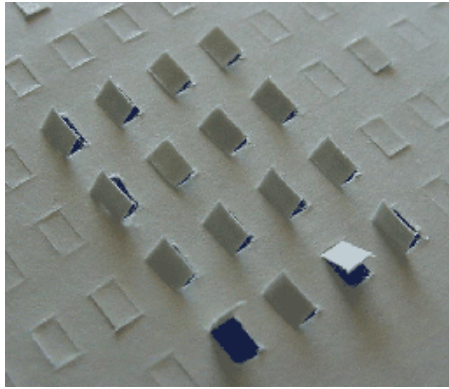
How did we get where we are?

The infamous butterfly ballot from the 2000 U.S. Presidential election:

The Florida ballot is a classic example of bad user interface design. (Computer software can suffer from such problems just as easily.)

http://www2.indystar.com/library/factfiles/gov/politics/election2000/img/prezrace/butterfly_large.jpg

Hanging chads & voter intent

Votomatic technology used in Florida was prone to paper jams. This led to hanging and dimpled chads, making it hard to determine *voter intent*, which provides the legal standard.

http://www.cs.uiowa.edu/~jones/cards/chad.html
http://www.pushback.com/justice/votefraud/DimpledChadPictures.html

Counting votes may not be easy

INSTRUCTIONS TO VOTERS
To vote, completely fill in the oval(s) next to your choice(s) like this: ⬤

Is this a legal vote?

- Courts would probably say so ...
- ... but op-scan readers might not count it.

Increasing demands that machine's interpretation match a human's.

Evaluating election technologies

Some general system-level goals for trustworthy elections:

- Need accurate determination of voter intent.
- Must preserve voter anonymity.
- Accessibility for disabled voters and non-native speakers.
- If possible, prevent overvoting (invalidates voter's ballot).
- If possible, prevent unintentional undervoting (voter confusion?).
- Easy to administer, even by under-trained poll workers.
- Transparently fair.

Lingering concerns about paper

Draft report on Voluntary Voting Systems Guidelines by the Security and Transparency Subcommittee for the Technical Guidelines Development Committee of the National Institute of Standards and Technology (NIST):

"The widespread adoption of voting systems incorporating paper did not seem to cause any widespread problems in the November 2006 elections. But, the use of paper in elections places more stress on (1) the capabilities of voting system technology, (2) of voters to verify their accuracy, and (3) of election workers to securely handle the ballots and accurately count them. Clearly, the needs of voters and election officials need to be addressed with improved and new technology. The STS believes that current paper-based approaches can be improved to be significantly more usable to voters and election officials ..."

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. http://vote.nist.gov/DraftWhitePaperOnSIinVVSG2007-20061120.pdf.

Research questions

Issues that arise from using paper ballots in elections:

- Accurate interpretation of marginal markings.
- Human cost, error rate, and bias in performing manual recounts.
- Failure modes in ballot imaging (e.g., paper jams).
- Systematic errors due to ballot layout (one candidate may be disadvantaged over another based on physical location on page).

Also keep in mind:

- U.S. Elections can be complex (10's to 100's of choices).
- Impact of "voter error" (e.g., improper markings, erasures).
- Potential for traditional ballot-box stuffing.
- Computer hackers attempting to manipulate the vote.

Connection to forms processing

Similarities to forms processing, but also some key differences:

- Much broader range of users (education level, literacy, etc.) than for traditional forms applications.
- Ballots must preserve a voter's anonymity.
- Demand to count votes and report results quickly.
- Elections are held infrequently, so voting equipment sits unused for long periods in storage.
- Poll workers often lack technical expertise.
- Maintaining chain-of-custody is a critical security requirement.
- No *financial* interest in making sure votes are counted accurately, but there is tremendous *public* interest.

BallotToolkit

Software components written in Tcl/Tk and runnable under both MS Windows and Linux. GUI logs user interactions (all events time-stamped) to facilitate user studies. Data interchange via XML-like file formats.

Provides support for:

- Ballot specification (locating targets, defining races and elections).
- Ballot ground-truthing (human interpretation of ballot markings).
- Synthesizing collections of marked ballots.
- Investigating "blind" auditing to eliminate human bias.
- Investigating homogeneous class display to facilitate recounts.

BallotTool software
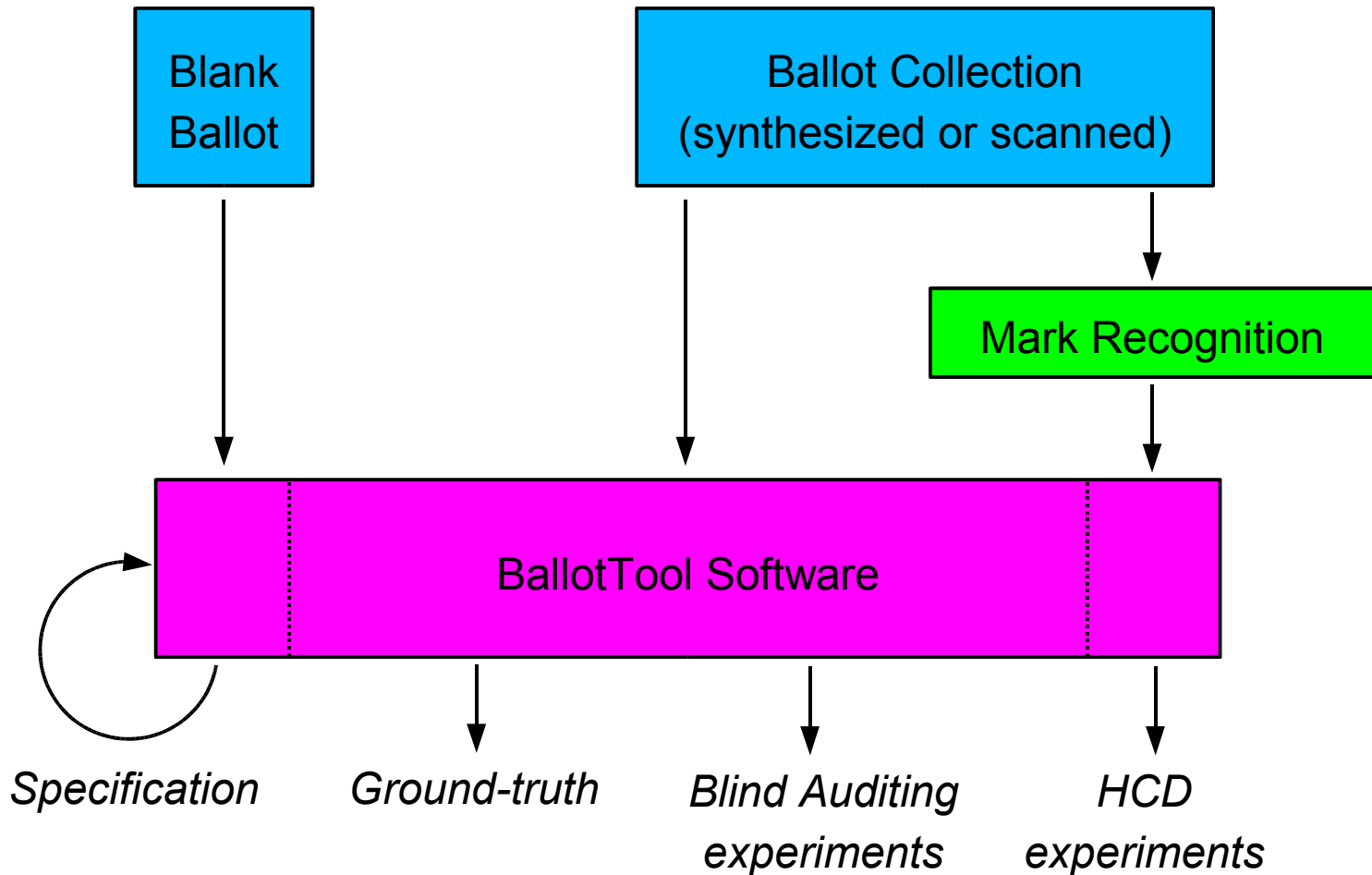
Blank Ballot

Ballot Collection (synthesized or scanned)

Mark Recognition

BallotTool Software

Specification

Ground-truth

Blind Auditing experiments

HCD experiments

A Document Analysis System for
Supporting Electronic Voting Research

Rensselaer

LEHIGH
UNIVERSITY™

BOISE STATE
UNIVERSITY

Lopresti • Slide 14
DAS Workshop • 9/19/08

BallotTool GUI

File format for specifying ballots

Blank ballot

Lehigh University / Muhlenberg College
2007 Presidential Election
NATIONAL SURVEY

Please chose only one answer for each question. Fill in the oval for the answer that you select for each question.

In determining your vote for president in the 2008 election please indicate how important the following issues will be:

	Very Important	Somewhat Important	Not Too Important	Not At All Important
The War in Iraq	O	O	O	O
Global Warming	O	O	O	O
The Economy	O	O	O	O
Terrorism	O	O	O	O
Taxes	O	O	O	O
Education	O	O	O	O
Health Care	O	O	O	O
Crime	O	O	O	O

Please identify your views on the following candidates for president:

	Very Favorable	Somewhat Favorable	Not Too Favorable	Not At All Favorable
Rudi Giuliani	O	O	O	O
Hillary Clinton	O	O	O	O
Barack Obama	O	O	O	O
Mitt Romney	O	O	O	O
John McCain	O	O	O	O

If the 2008 presidential election was being held today and the candidates were Democrat Hillary Clinton and Republican Rudy Giuliani, who would you vote for?

Hillary Clinton   O
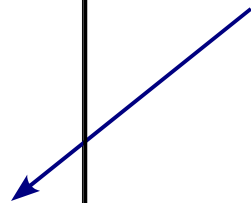Rudi Giuliani   O

If the 2008 presidential election was being held today and the candidates were Democrat Barack Obama and Republican Rudy Giuliani, who would you vote for?

Barack Obama   O
Rudi Giuliani   O

Associated specification

```
<election ID="election001" Election="Lehigh-Muhlenberg Survey"
        bb_x1="10" bb_y1="10" bb_x2="2542" bb_y2="3290">
<race ID="race001" Race="The War in Iraq" VoteFor="1"
        bb_x1="350" bb_y1="1050" bb_x2="800" bb_y2="1105">
<candidate ID="cand001" Candidate="Very Important"
        bbl_x1="830" bbl_y1="890" bbl_x2="1060" bbl_y2="1020"
        bbt_x1="900" bbt_y1="1050" bbt_x2="990" bbt_y2="1105">
<candidate ID="cand002" Candidate="Somewhat Important"
        bbl_x1="1130" bbl_y1="890" bbl_x2="1360" bbl_y2="1020"
        bbt_x1="1200" bbt_y1="1050" bbt_x2="1290" bbt_y2="1105">
<candidate ID="cand003" Candidate="Not Too Important"
        bbl_x1="1430" bbl_y1="890" bbl_x2="1660" bbl_y2="1020"
        bbt_x1="1500" bbt_y1="1050" bbt_x2="1590" bbt_y2="1105">
<candidate ID="cand004" Candidate="Not At All Important"
        bbl_x1="1730" bbl_y1="890" bbl_x2="1960" bbl_y2="1020"
        bbt_x1="1800" bbt_y1="1050" bbt_x2="1890" bbt_y2="1105">
</race>
<race ID="race002" Race="Global Warming" VoteFor="1"
        bb_x1="350" bb_y1="1110" bb_x2="800" bb_y2="1165">
<candidate ID="cand001" Candidate="Very Important"
        bbl_x1="830" bbl_y1="890" bbl_x2="1060" bbl_y2="1020"
        bbt_x1="900" bbt_y1="1110" bbt_x2="990" bbt_y2="1165">
'''
```

BallotTool GUI for blind auditing
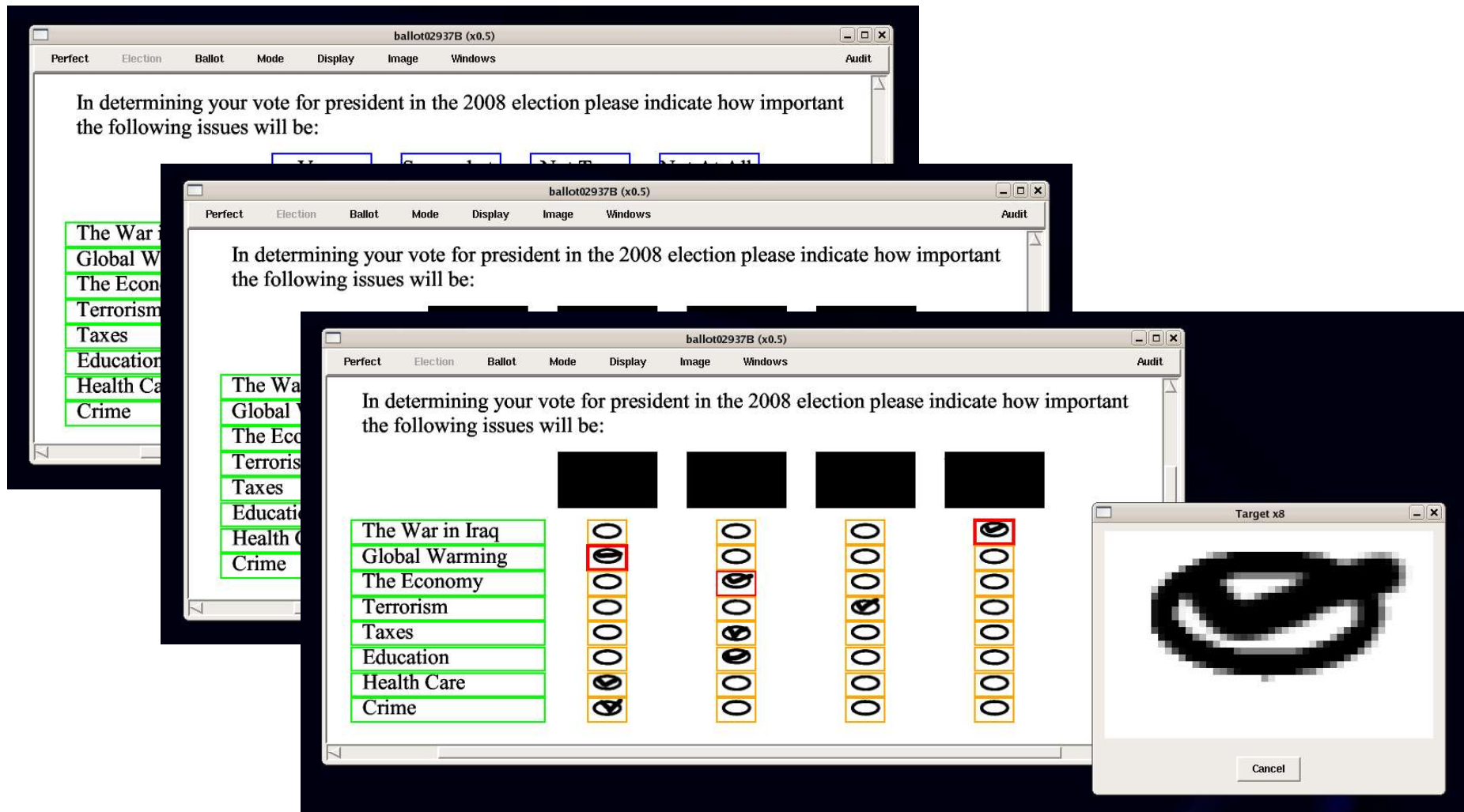
BallotTool GUI for HCD

BallotGen software

Blank Ballot

Mark Library

Election Specification
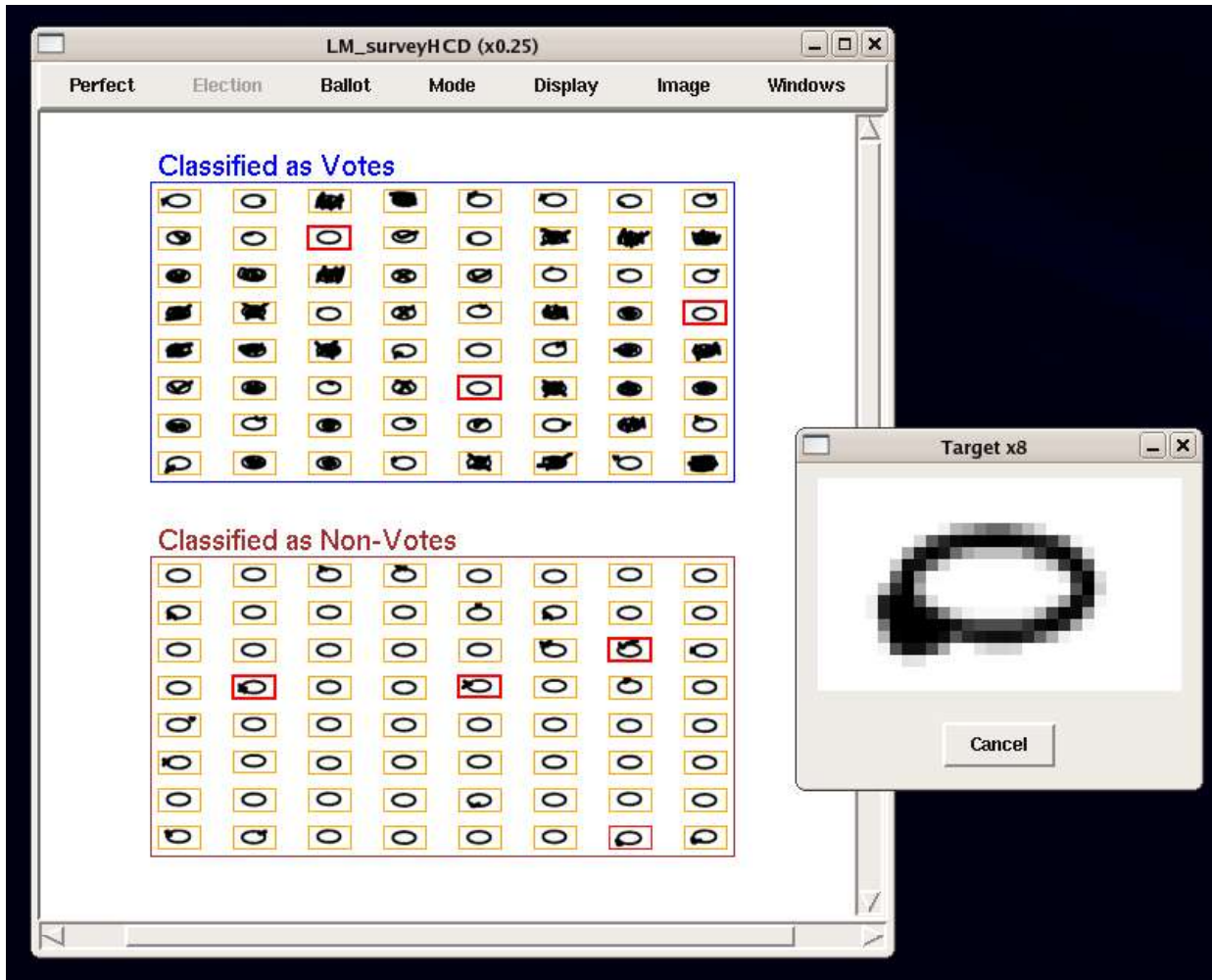
BallotGen Software

Synthetic Ballot Collection (PDF, TIF)

Print / Scan

Machine interpretation

Human interpretation
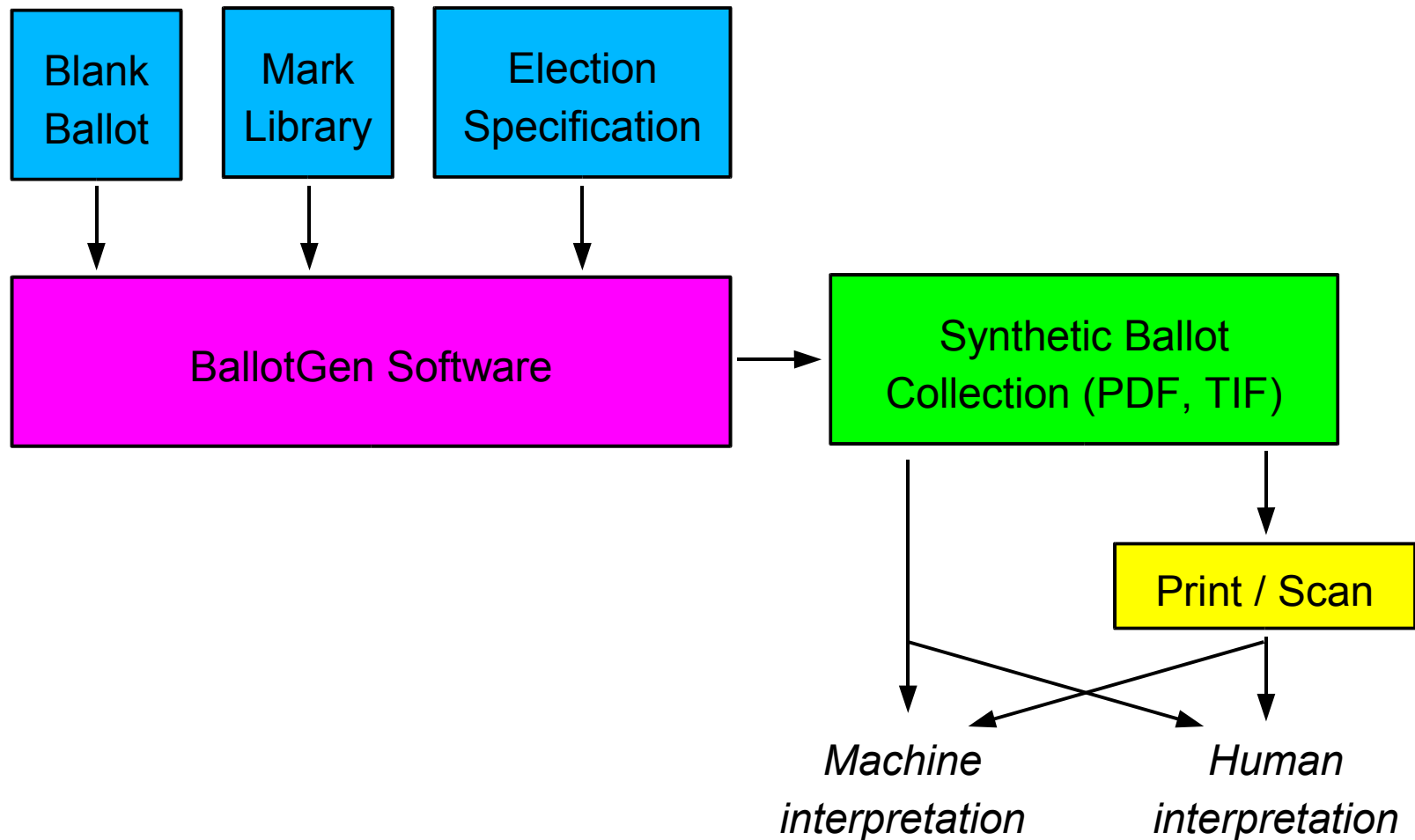
Synthesizing ballots

Two paradigms for injecting marks on blank ballot substrate:

- Extract and place pre-marked targets on image.
- Transform and overlay marks with transparent backgrounds.

In latter case, we can:

- Adjust x- and y-displacement from target center.
- Scale x- and y-dimensions independently.
- Rotate mark by a random amount.
- Re-map grayscale or color of mark.

Pre-marked targets

Synthesizing ballots

Simulated Lehigh-Muhlenberg 2008 Presidential Election survey. Synthesized using marks that are randomly chosen and placed (some intentionally "marginal").

Transformed and overlayed marks

Synthesizing ballot collections

Step 4: define election (# of ballots, undervote and overvote rates, etc.)
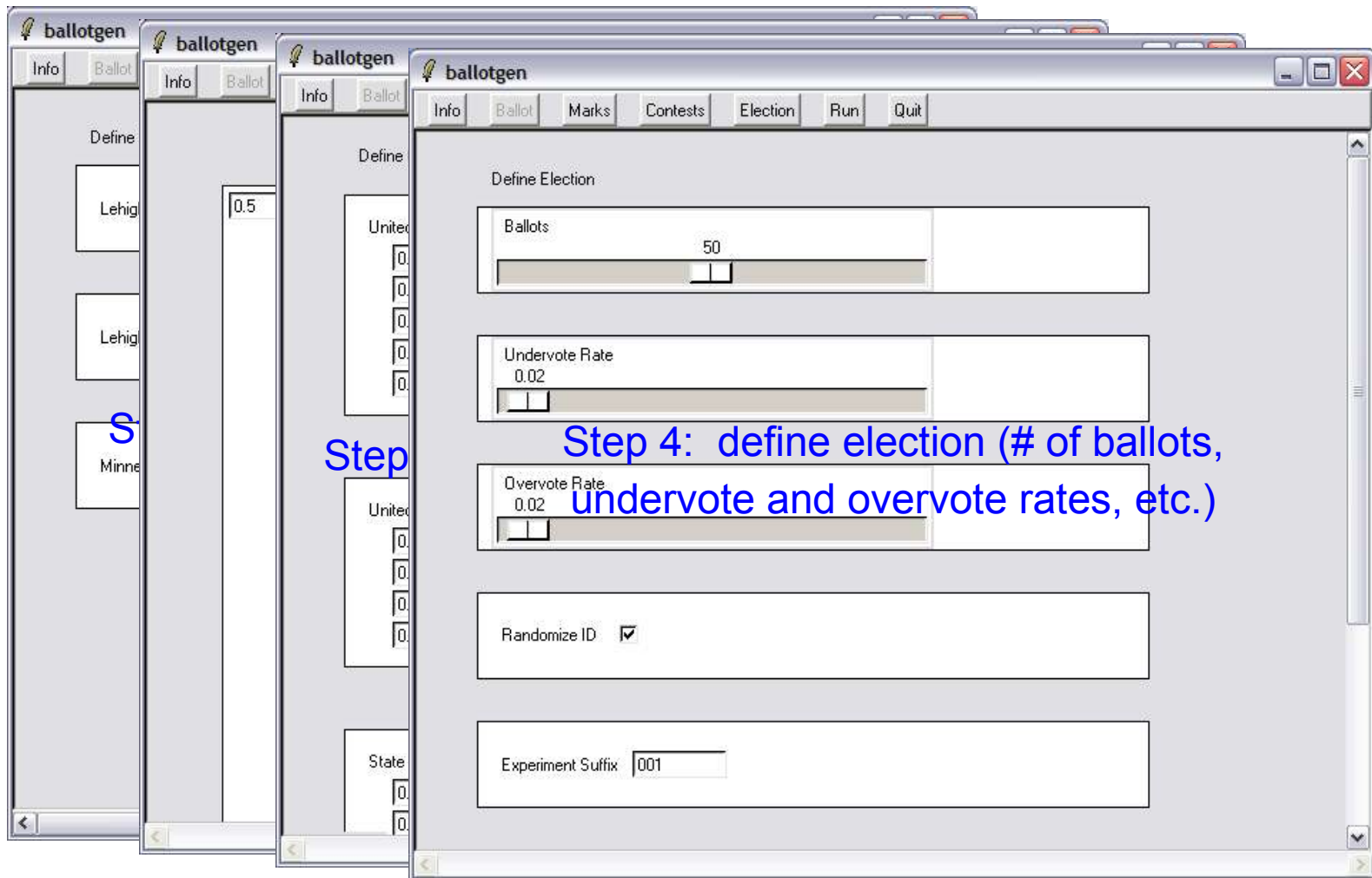
Summary

Status:

- Prototype nearly complete – blind auditing and HCD experiments will commence soon in collaboration with social scientist colleagues. Results to be presented in future papers.

Conclusions:

- Paper ballot processing provides an opportunity to apply document analysis research to a timely and important problem.

- Upon reflection, a number of other ideas will come to mind. E.g., style-based recognition for interpreting marginal markings.

Whole-ballot recognition

Stray mark?

Valid vote?

BASEBALL HALL OF FAME
(vote for no more than 5)

- Ty Cobb
- Rogers Hornsby
- Walter Johnson
- Nap Lajoie
- Christy Mathewson
- Babe Ruth
- Tris Speaker
- Honus Wagner
- Cy Young

BASEBALL HALL OF FAME
(vote for no more than 5)

- Ty Cobb
- Rogers Hornsby
- Walter Johnson
- Nap Lajoie
- Christy Mathewson
- Babe Ruth
- Tris Speaker
- Honus Wagner
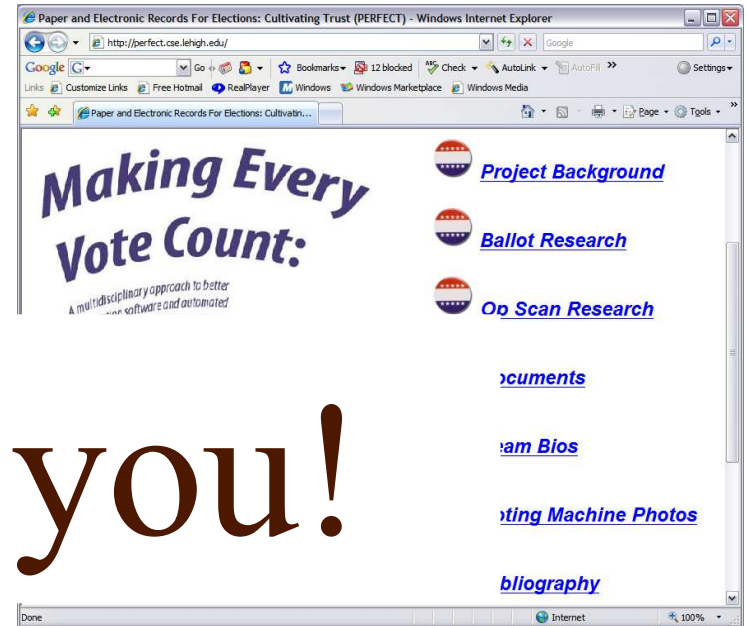- Cy Young

$\Rightarrow$ Capture voter intent via style-based techniques.

Paper and Electronic Records
for Elections: Cultivating Trust

Thank you!

*A Document Analysis System for
Supporting Electronic Voting Research*

Rensselaer   LEHIGH   BOISE STATE
UNIVERSITY

*Lopresti • Slide 27
DAS Workshop • 9/19/08*