Document Analysis Issues in Reading Optical Scan Ballots

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

Portion of an op-scan ballot

In a nutshell

Reading optical scan ballots presents an interesting and important challenge. In addition to our helping solve it, what can be learned?

- Task is extremely well specified (codified into law) and cannot be altered just to make pattern recognition problems easier.

- Like most pattern recognition problems, 100% accuracy may not be attainable, but in this case it is at least a reasonable goal.

Primary contribution of this work:

- Assembling and making openly available a large collection of real hand-marked ballots from a major election. (A "first"?)

- Proposing a paradigm for ground-truthing the collection.

Ballot reading vs. 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.
- Requirements legally mandated – cannot be circumvented.
- 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.
- No *financial* interest in making sure votes are counted accurately, but there is tremendous *public* interest.

Misstep #1: Florida circa 2000
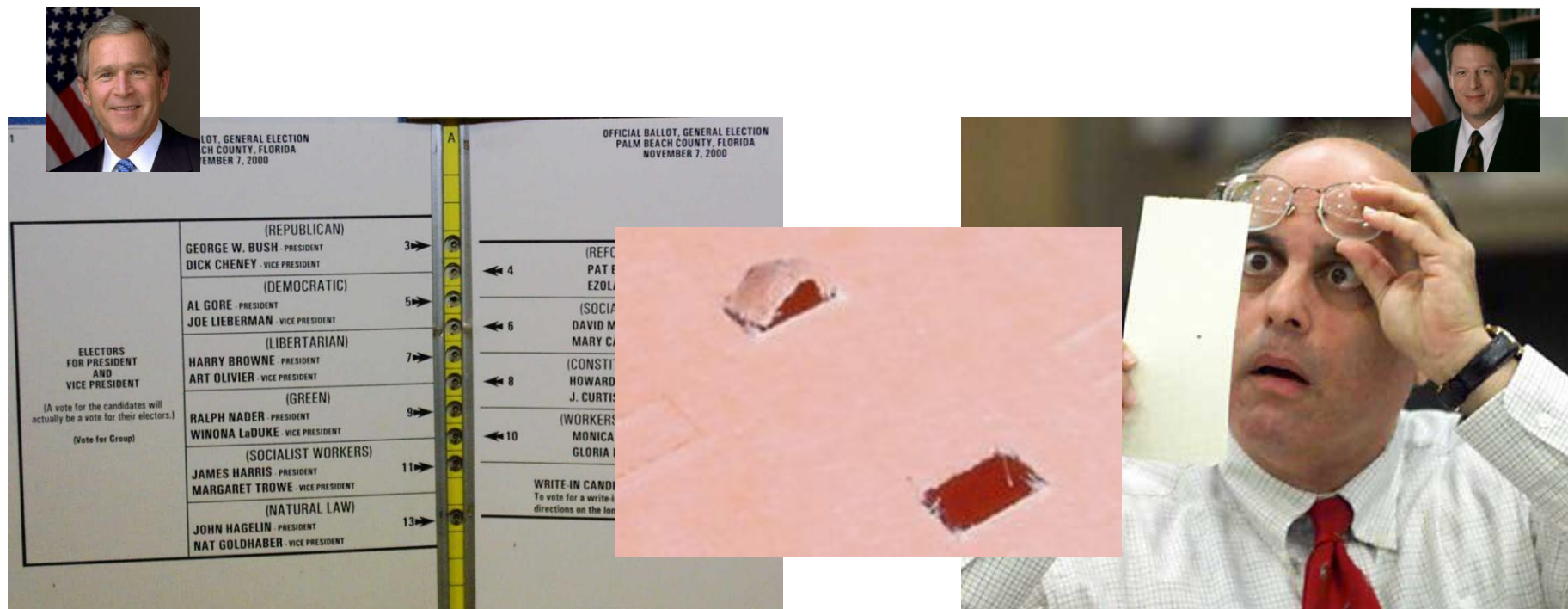
Bad ballot design combined with votomatic technology prone to paper jams. This led to hanging and dimpled chads, making it hard to determine *voter intent*, which provides the legal standard.

Misstep #2: Rush to DREs …
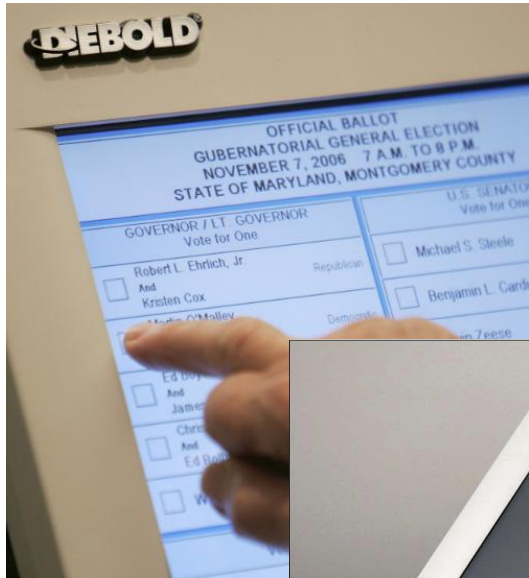
… Misstep #2: Rush to DREs

Security Analysis of the Diebold AccuVote-TS Voting Machine

Electronic Voting Systems: the Good, the Bad, and the Stupid

SECURITY ALERT: July 4, 2005
Critical Security Issues with Diebold Optical Scan Design

Security Assessment of the Diebold Optical Scan Voting Terminal

Pennsylvania voters: trust but verify
Poll finds most want ballot verification

Electronic Voting System Usability Issues

THE MACHINERY OF DEMOCRACY:
PROTECTING ELECTIONS
IN AN ELECTRONIC WORLD

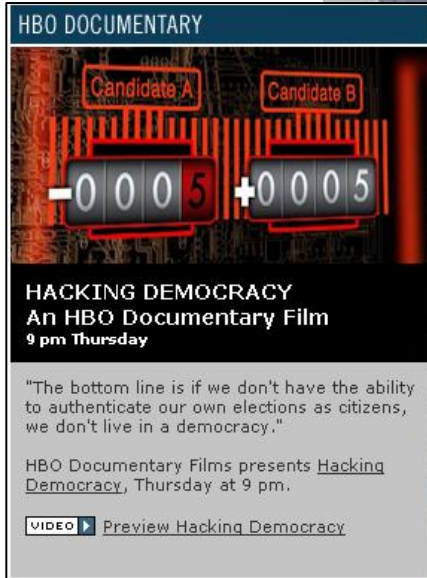Hack-a-Vote: Security Issues with Electronic Voting Systems

Analysis of an Electronic Voting System

Privacy Issues in an Electronic Voting Machine

SECURITY ALERT: May 11, 2006
Critical Security Issues with Diebold TSx

Trusted Agent Report
Diebold AccuVote-TS Voting System

HBO DOCUMENTARY

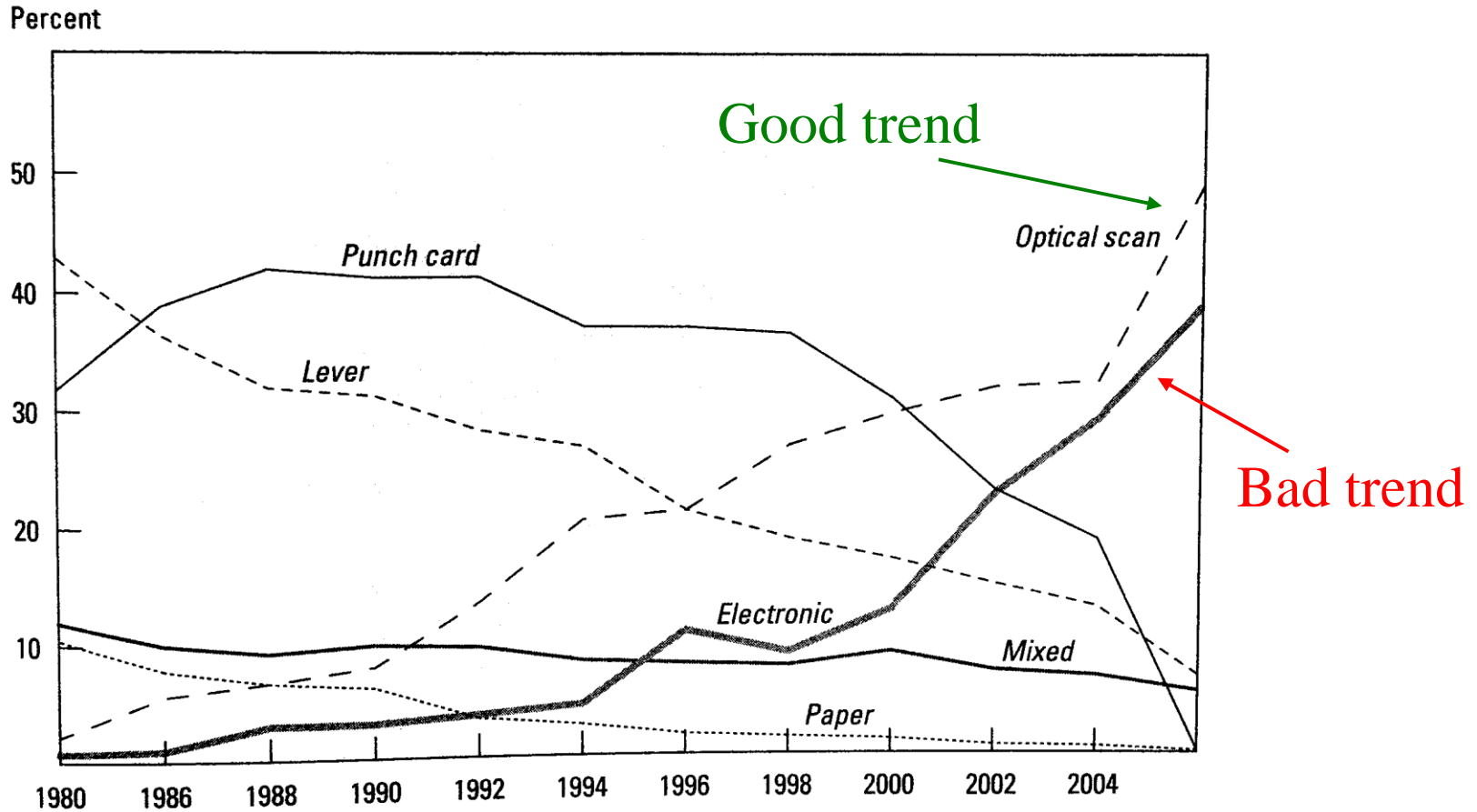Candidate A   Candidate B
- 0 0 0 5   + 0 0 0 5

HACKING DEMOCRACY
An HBO Documentary Film
9 pm Thursday

"The bottom line is if we don't have the ability to authenticate our own elections as citizens, we don't live in a democracy."

HBO Documentary Films presents Hacking Democracy, Thursday at 9 pm.
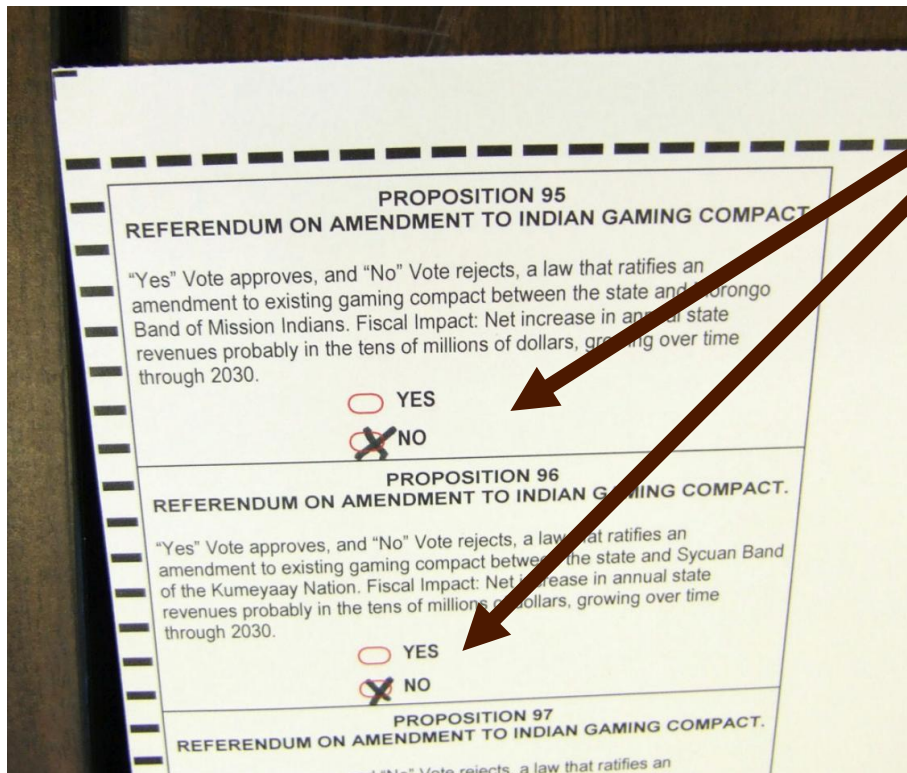
VIDEO ▶ Preview Hacking Democracy

Voting system use in the U.S.

Counting op-scan votes is not so easy

Real ballot from an election in CA:

One of these votes was counted correctly by the op-scan equipment, the other wasn't.

Note: this does <u>not</u> mean voting on paper ballots is bad, just (1) manual audits should be mandatory, and (2) more research is needed.

"Improving California's 1% Manual Tally Procedure," Joseph Lorenzo Hall, UC Berkeley School of Information, EVT Workshop 2008.

It's all about the data

To do research, we need data. We could:

- Hire students to fill out sample ballots.
- Generate synthetic data (see examples below from our DAS 2008 paper).

Fast forward to November 2008

While useful, neither of these alternatives is completely satisfying.

What we would like to have is actual ballots from a real election. Even better, the ballots would be from an important election where the voter markings present serious pattern recognition challenges.

Extremely close U.S. Senate race in State of Minnesota: six days after election, unofficial results showed Republican Norm Coleman leading Democratic challenger Al Franken by 206 votes out of nearly 3 million cast, a difference of less than 0.01%.

Document Analysis Issues in Reading Optical Scan Ballots

Rensselaer

LEHIGH UNIVERSITY

BOISE STATE UNIVERSITY

Lopresti • Slide 11
DAS Workshop • 6/10/10

Fortunate series of events

- Minnesota uses op-scan ballots.

- Closeness of election triggers a manual recount.

- Both sides are allowed to challenge validity of "questionable" ballots.

- Openness laws make challenged ballots a matter of public record.

- Ballot images made available on MN public radio website.

- PDF files contain 300 dpi TIF images!

http://minnesota.publicradio.org/features/2008/11/19_challenged_ballots/

Minnesota Statutes

Remember that the guiding principle is *voter intent*. Here are a few key points to keep in mind when interpreting ballot markings:

- "A ballot shall not be rejected for a technical error that does not make it impossible to determine the voter's intent."

- "If a mark (X) is made out of its proper place, but so near a name or space as to indicate clearly the voter's intent, the vote shall be counted."

- "Misspelling or abbreviations of the names of write-in candidates shall be disregarded if the individual for whom the vote was intended can be clearly ascertained from the ballot."

https://www.revisor.mn.gov/statutes/?id=204C.22

Minnesota Statutes

… and …

- "If a voter uniformly uses a mark other than (X) which clearly indicates an intent to mark a name or to mark yes or no on a question, and the voter does not use (X) anywhere else on the ballot, a vote shall be counted for each candidate or response to a question marked.
  If a voter uses two or more distinct marks, such as (X) and some other mark, a vote shall be counted for each candidate or response to a question marked, unless the ballot is marked by distinguishing characteristics that make the entire ballot defective …"

https://www.revisor.mn.gov/statutes/?id=204C.22

Document Analysis Issues in Reading Optical Scan Ballots

Rensselaer

LEHIGH UNIVERSITY

BOISE STATE UNIVERSITY

Lopresti ● Slide 14
DAS Workshop ● 6/10/10

Minnesota Statutes

… and …

- "If the names of two candidates have been marked, and an attempt has been made to erase or obliterate one of the marks, a vote shall be counted for the remaining marked candidate."

- "A ballot shall not be rejected merely because it is slightly soiled or defaced."

- "If a ballot is marked by distinguishing characteristics in a manner making it evident that the voter intended to identify the ballot, the entire ballot is defective."

Goal here is to prevent coercion or vote selling.

https://www.revisor.mn.gov/statutes/?id=204C.22

Document Analysis Issues in Reading Optical Scan Ballots

Rensselaer

LEHIGH UNIVERSITY™

BOISE STATE UNIVERSITY

Lopresti • Slide 15
DAS Workshop • 6/10/10

Challenged ballots: you be the judge

Who gets vote? Public opinion:

- Norm Coleman: 63% (7,626 votes)
- Al Franken: 4% (474 votes)
- Nobody: 33% (4,050 votes)

Challenged ballots: you be the judge

Vote for Franken? Public opinion:

- Yes: 92% (11,069 votes)
- No: 8% (1,012 votes)

Document Analysis Issues in
Reading Optical Scan Ballots

Rensselaer

LEHIGH
UNIVERSITY

BOISE STATE
UNIVERSITY

Lopresti • Slide 17
DAS Workshop • 6/10/10

Challenged ballots: you be the judge

Vote for Franken? Public opinion:

- Yes: 96% (11,250 votes)
- No: 4% (452 votes)

Document Analysis Issues in Reading Optical Scan Ballots

Rensselaer

LEHIGH UNIVERSITY

BOISE STATE UNIVERSITY

Lopresti • Slide 18
DAS Workshop • 6/10/10

Challenged ballots: you be the judge
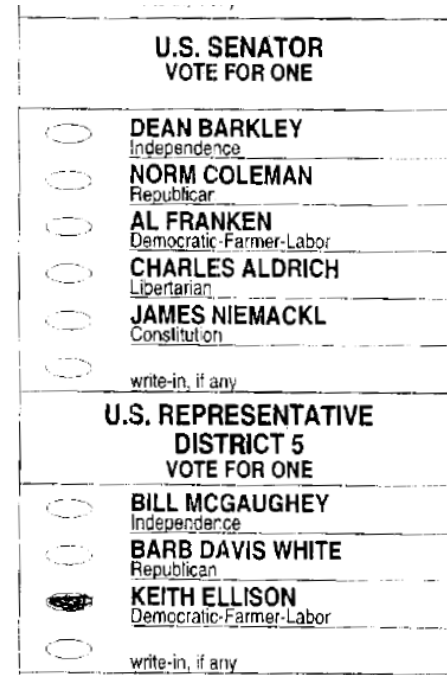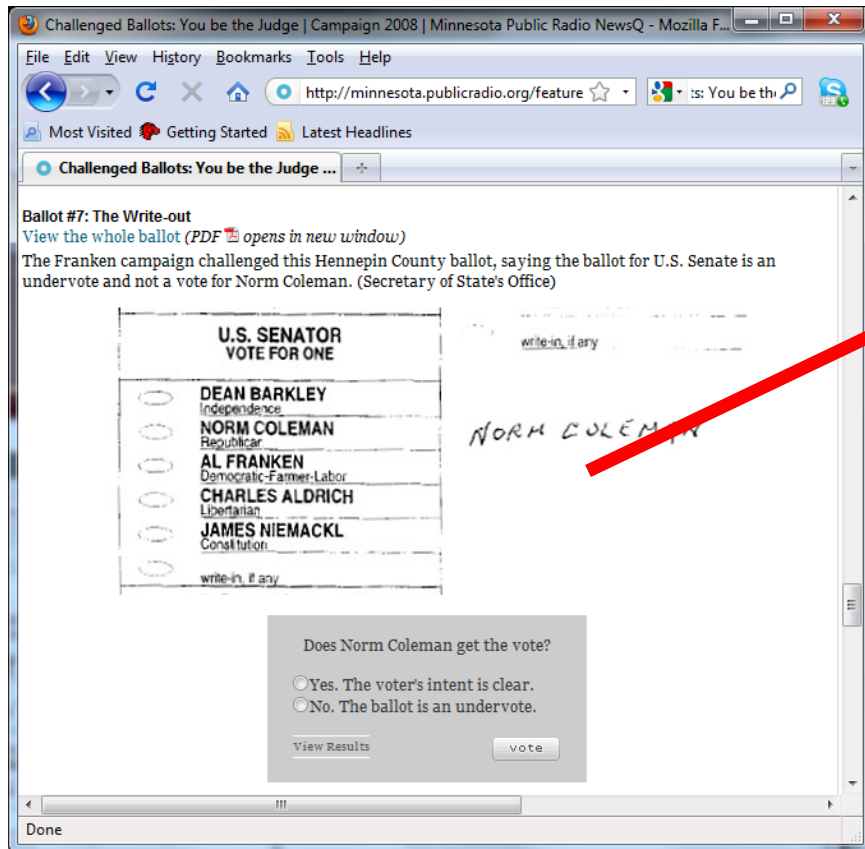
Vote for Coleman? Public opinion:

- Yes: 54% (6,080 votes)
- No: 46% (5,203 votes)

MN Challenged Ballot Collection

How the ballot collection was generated and harvested:

- Ballots photocopied and originals stored in a secure location.

- Copies scanned to PDF using auto-feeder flatbed scanner.

- Ballot was two-sided, with both sides scanned simultaneously.

- We wrote a simple web "crawler" that automatically downloaded all the files and extracted TIF images from PDF.

- A total of 6,737 ballots in the set.

- Examination of the TIF suggests that ballots were scanned at 300 dpi bitonal, and that lossy compression was never used.

- Hence, they form an ideal dataset for research purposes.

Document Analysis Issues in Reading Optical Scan Ballots

RENSSELAER

LEHIGH UNIVERSITY

BOISE STATE UNIVERSITY

Lopresti • Slide 20
DAS Workshop • 6/10/10

Minnesota ballot front and back

More about the collection

Reason for challenge is recorded in stamp added to each ballot:

U.S. Senator
Precinct: *Shamrock*
Challenged ballot #: *1*
Challenged by COLEMAN
Challenge reason: *Red stray mark on front of ballot*

- Challenges as a result of manual recount – not machine rejects.

- Ballot is challenged for one race – others may be "normal."

- Challenges are extremely aggressive (both sides want to win).

Franken ultimately declared winner by 312 votes: machine-reading got the results of the election wrong.

Ground-truthing protocol

Guiding principles:

- "Ground-truthing" can refer to identifying raw markings on ballot and/or interpreting their meaning ("vote" / "no vote").

- We have built a GUI tool that captures truth at both levels.

- Ideally, a system built for reading op-scan ballots should replicate same understanding of voter intent possessed by a knowledgeable human judge.

- Legitimate differences of opinion exist, so "truth" is relative.

- Users provided with 15-page guideline, as well as MN statutes.

What here generalizes to other document analysis applications?

BallotTool GUI

Document Analysis Issues in
Reading Optical Scan Ballots

Rensselaer

LEHIGH
UNIVERSITY

BOISE STATE
UNIVERSITY

Lopresti • Slide 24
DAS Workshop • 6/10/10

Ground-truth

Physical level:

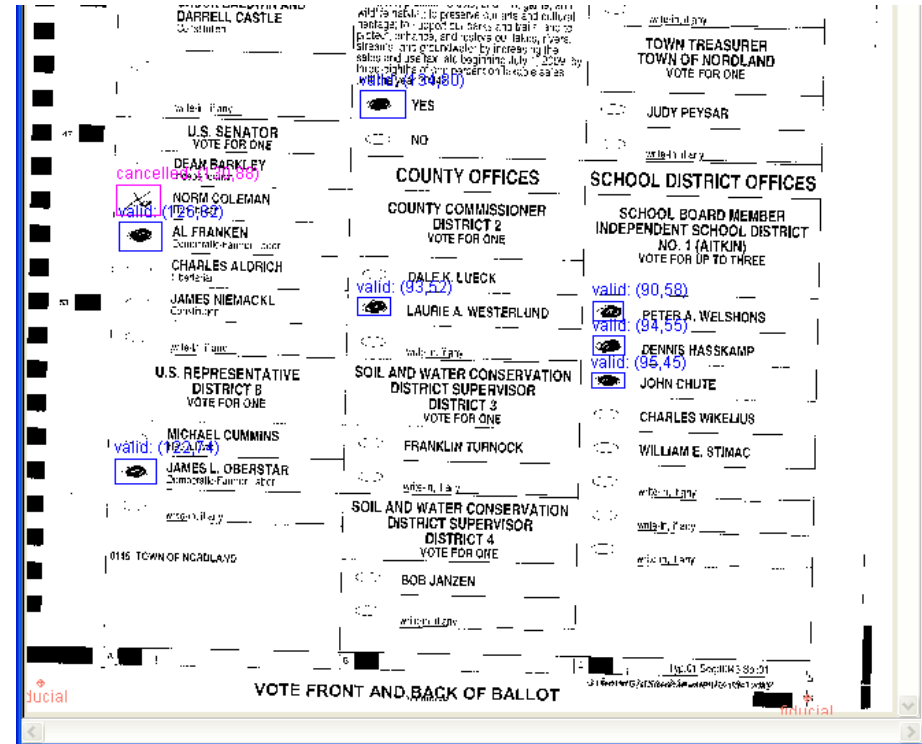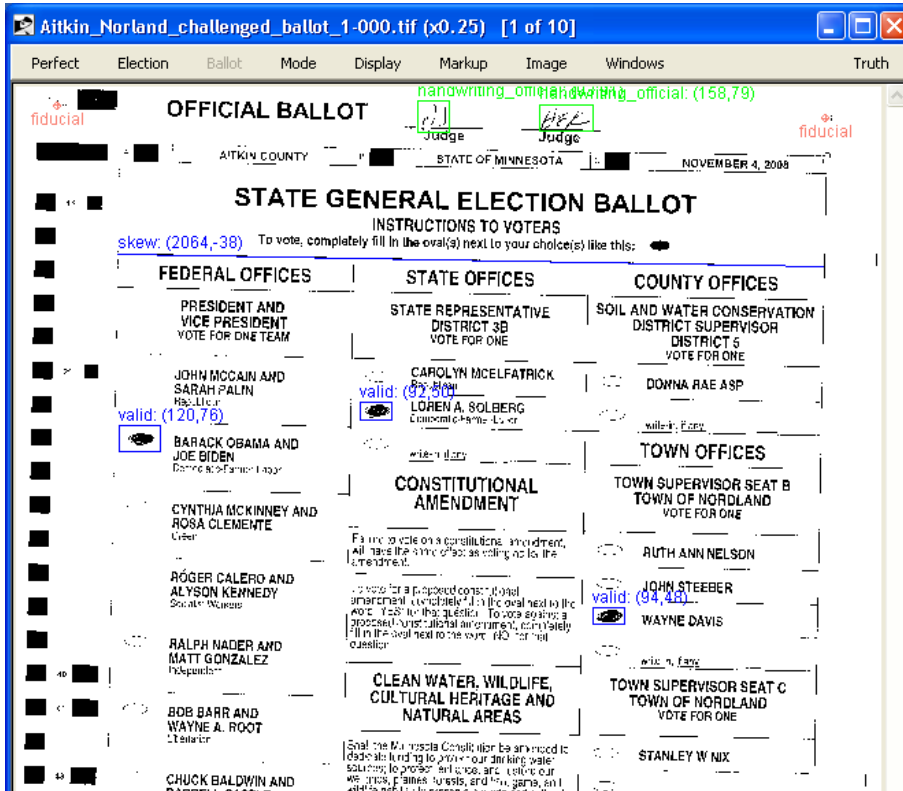- Each target annotated with bounding box as valid mark, stray mark, cancelled vote, or no vote.

- Marking style is recorded as filled oval, partially filled oval, ex (X), or check mark.

- Other marks on ballot face annotated (challenge stamp, bleed-through, voter or official handwriting, fiducial, skew line, etc.).
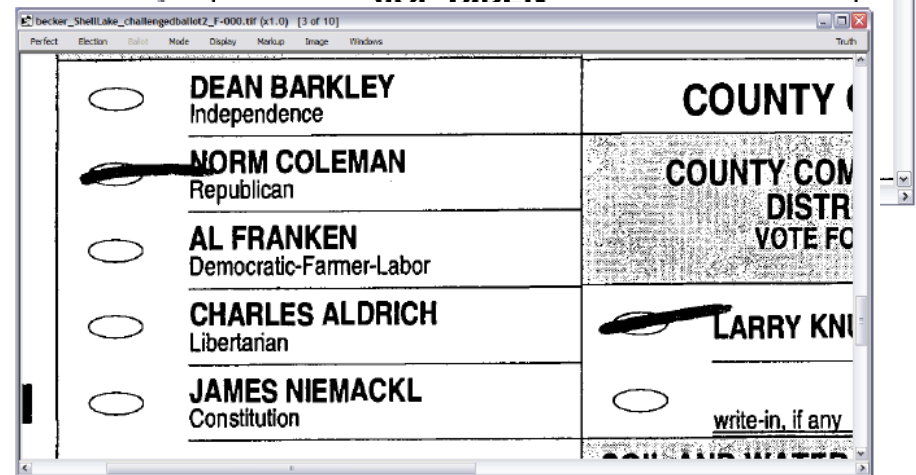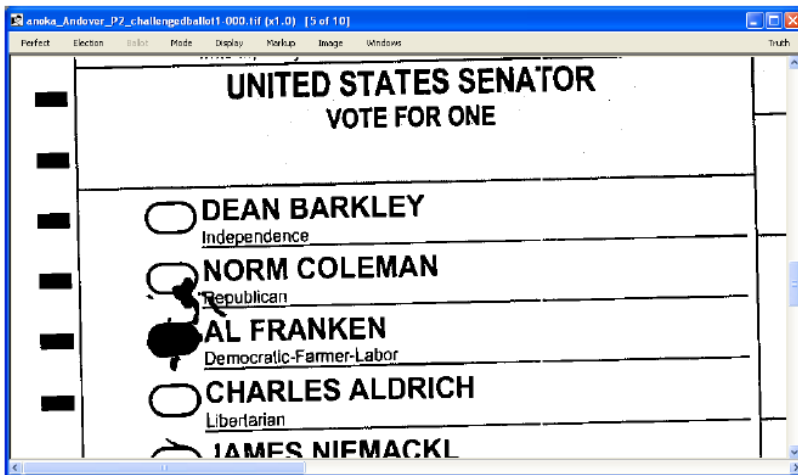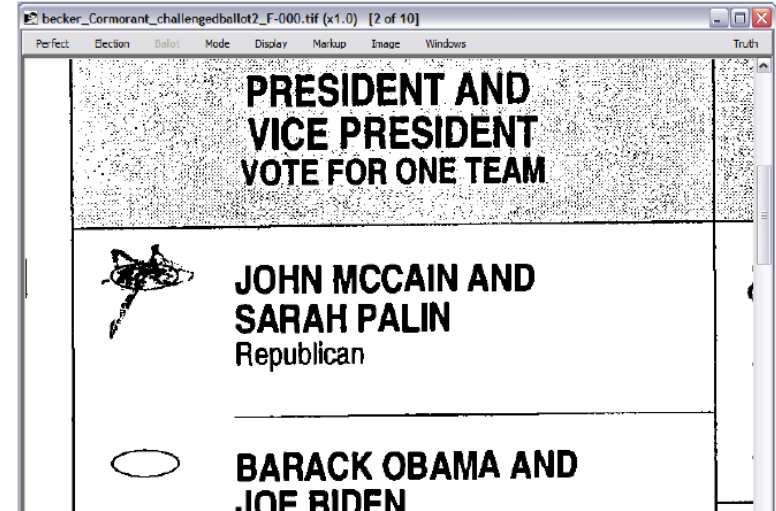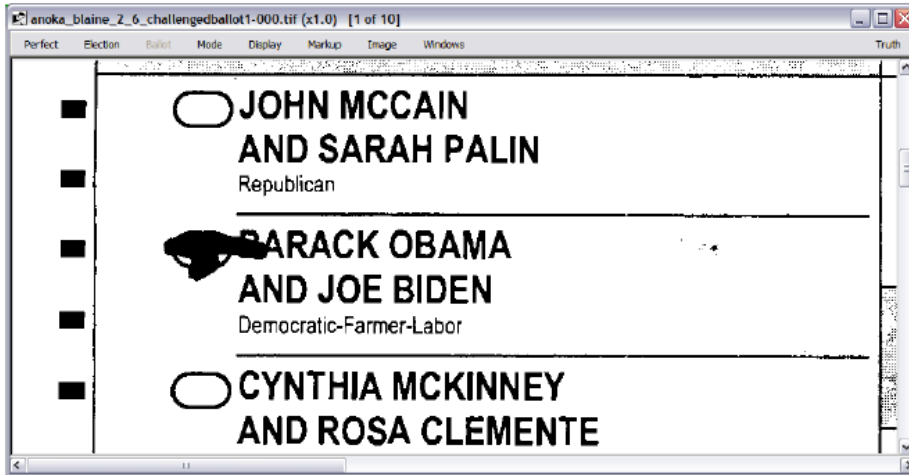
Logical level:

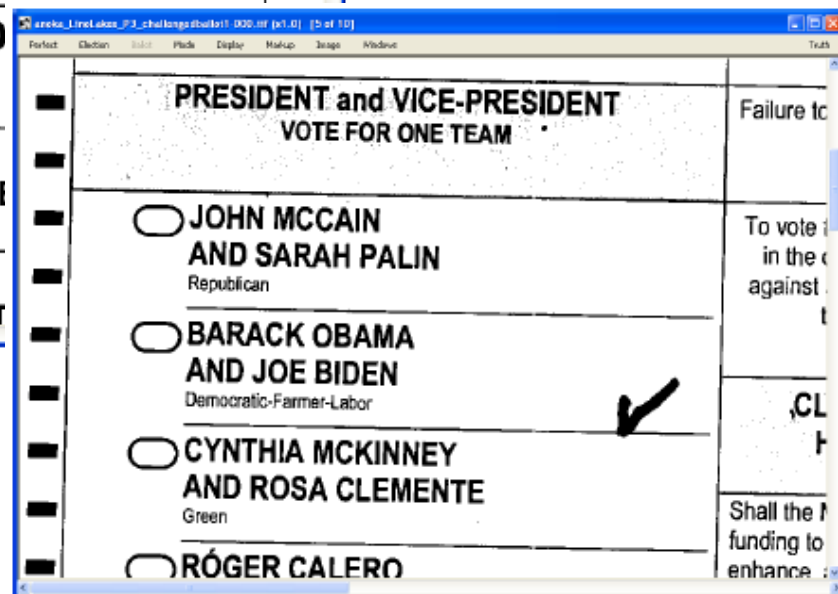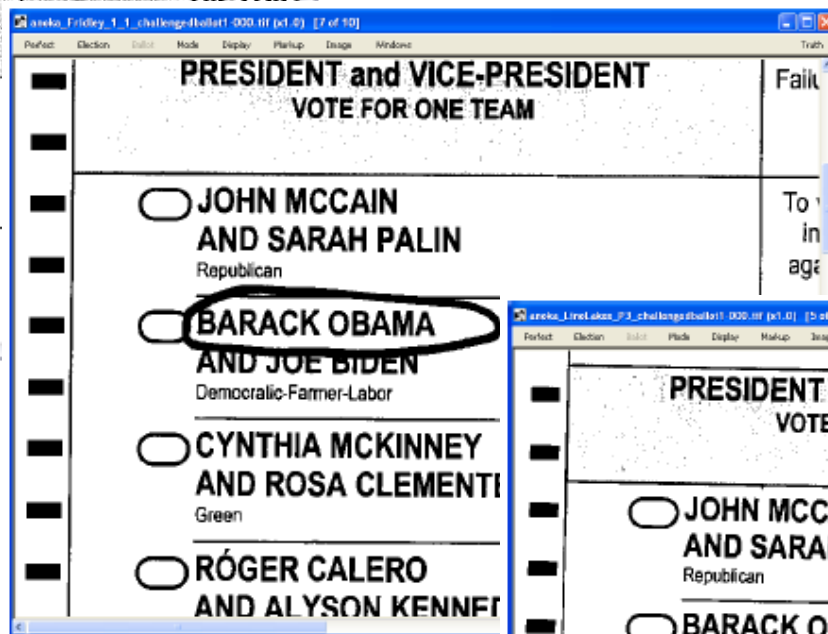Note two are not equivalent.

- User specifies who gets legal vote(s) in each race (i.e., intent).

Example of a Fully Annotated Ballot

Document Analysis Issues in
Reading Optical Scan Ballots

Rensselaer

LEHIGH
UNIVERSITY

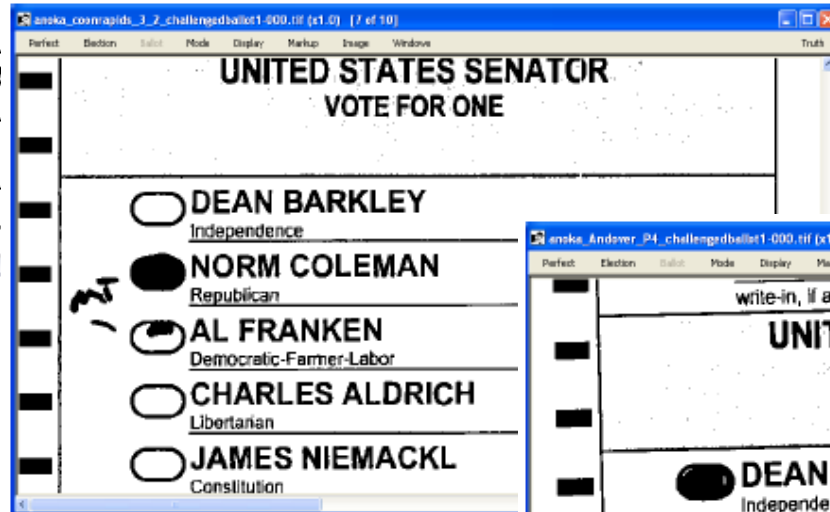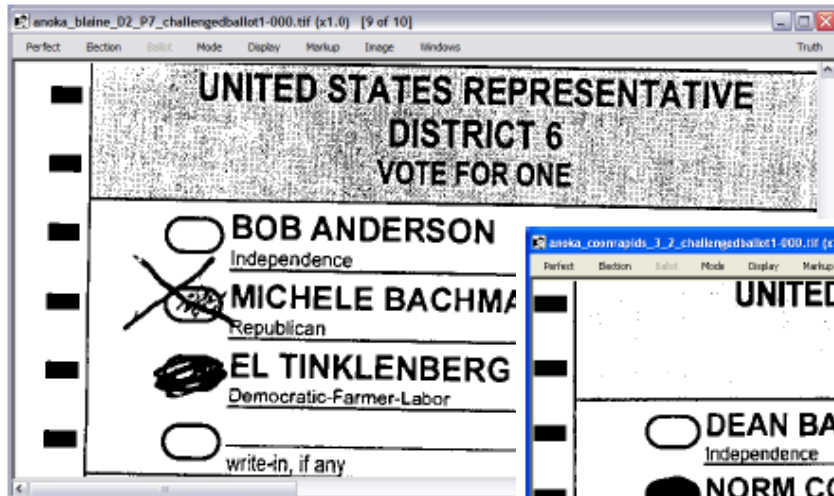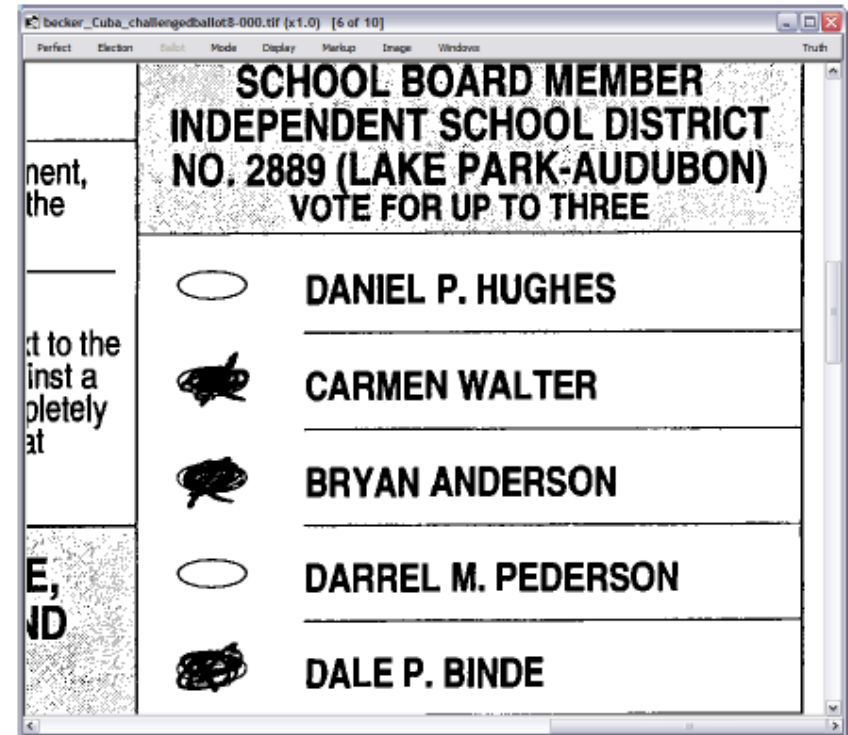BOISE STATE
UNIVERSITY

Lopresti • *Slide 26*
DAS Workshop • *6/10/10*

Sloppy-but-valid marks

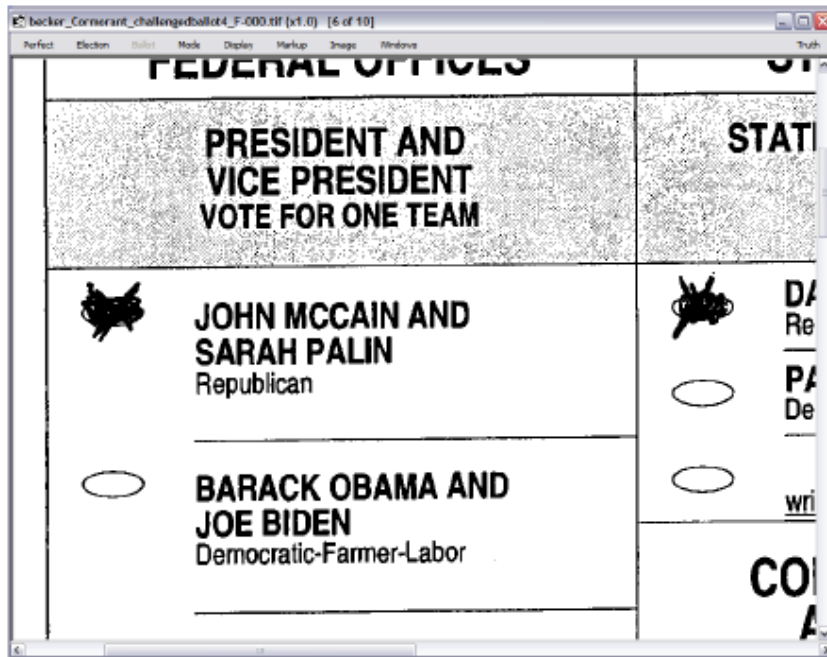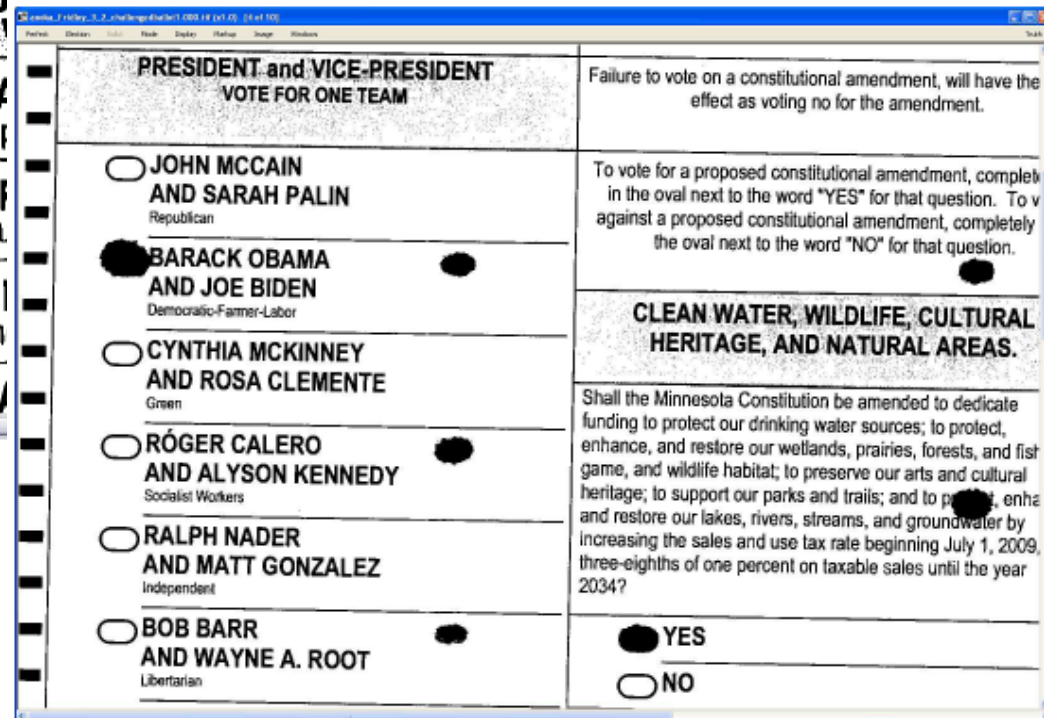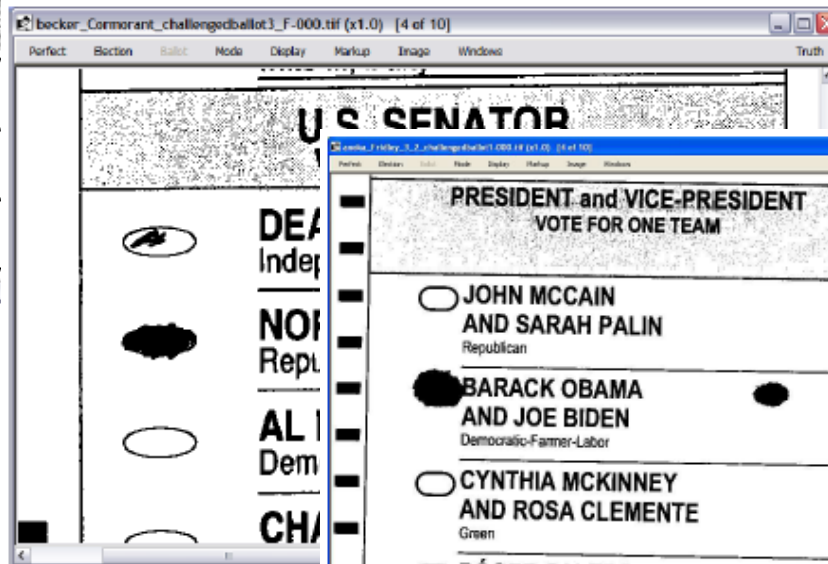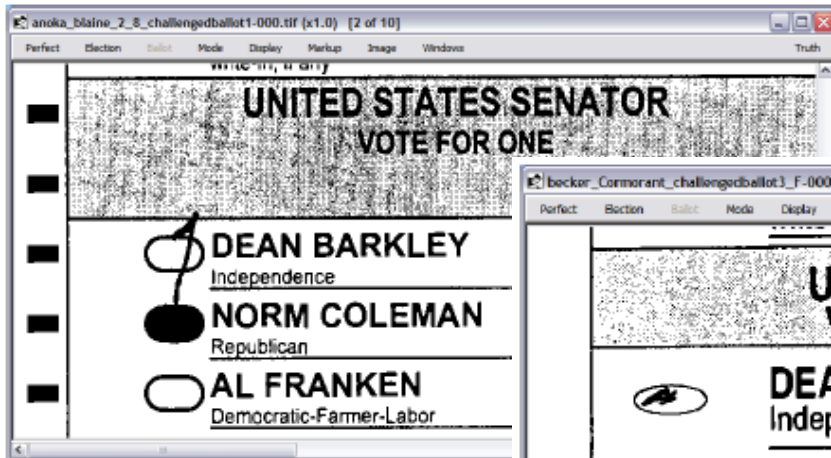# Non-confirming marking styles
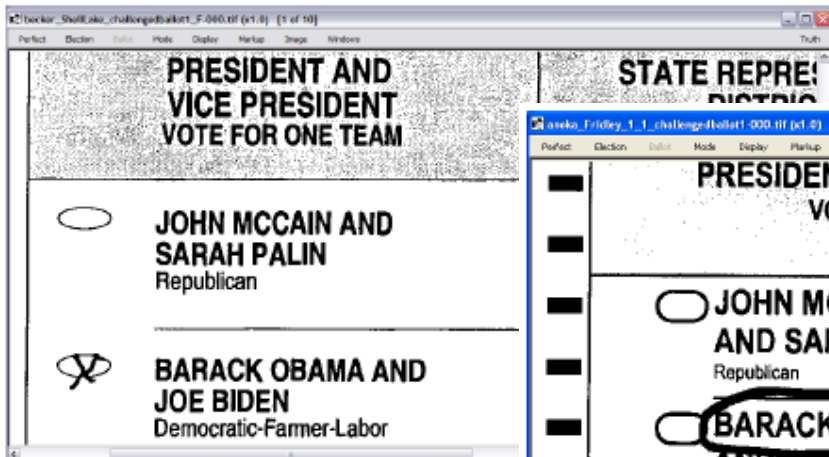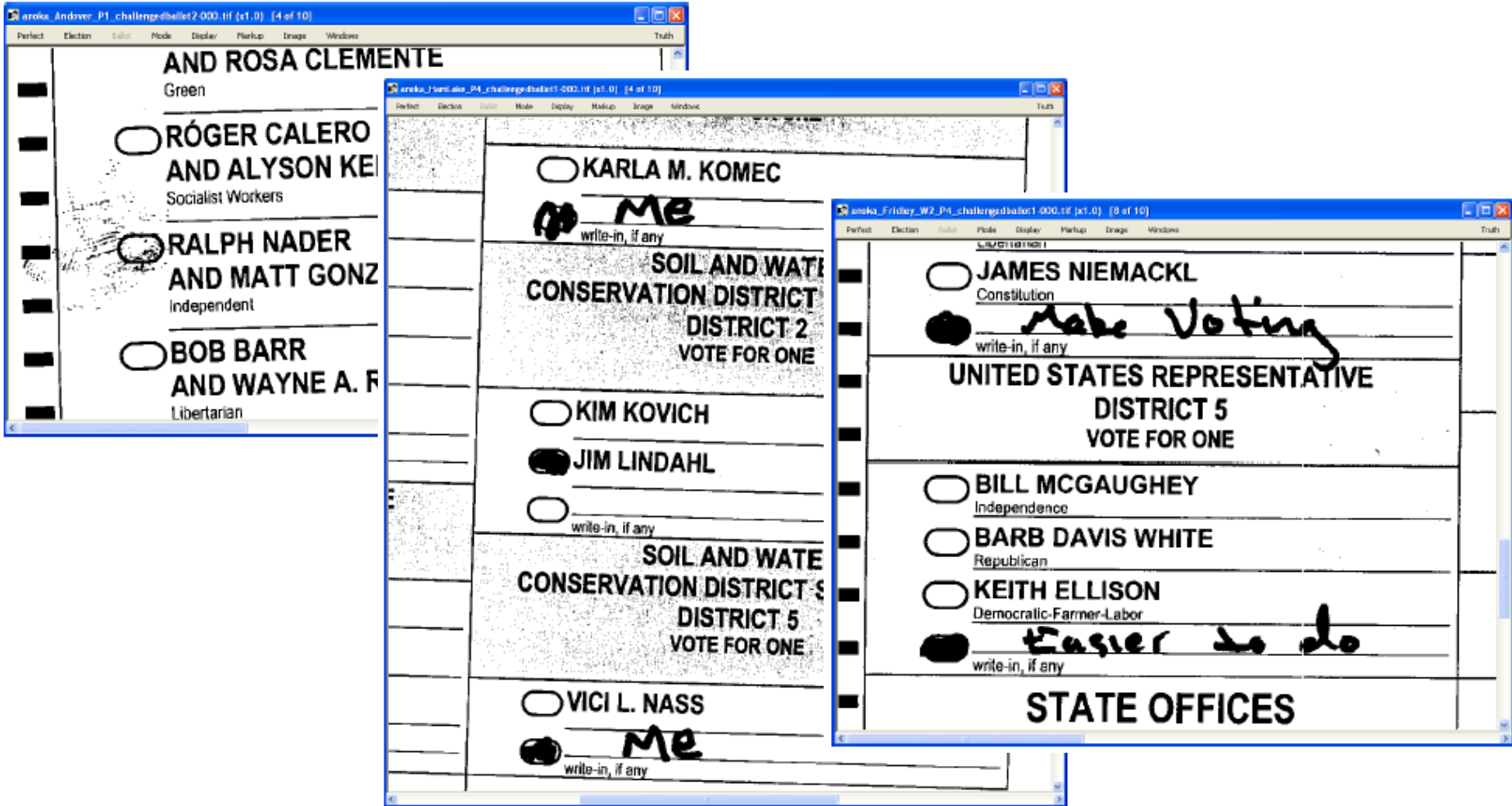
Attempts to cancel a vote
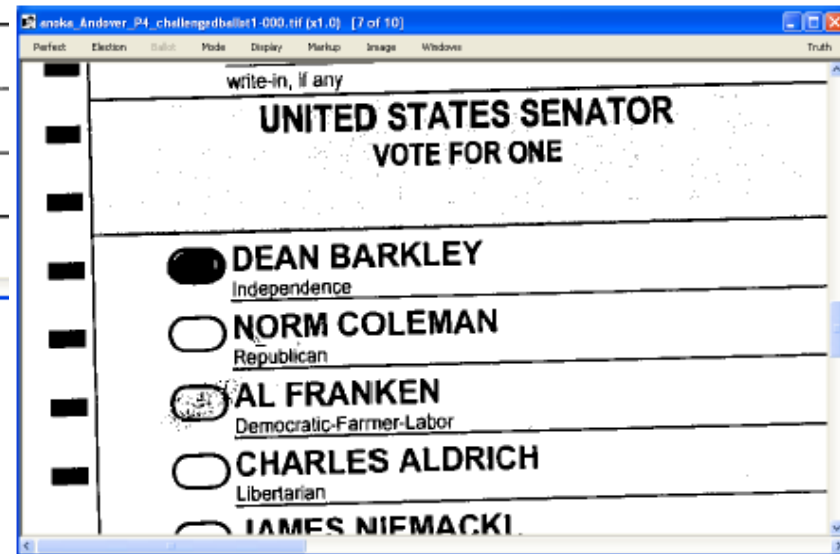
Intended votes that look cancelled

Stray marks and bleedthrough

Invalidating markings

Why isn't this an easy problem?

After all, ballots are just a simple type of form. We must read the votes correctly, but we aren't expected to recognize write-in names.

Can't we just push up reject rate until the accuracy reaches 100%?

Remember, we can't change rules in ways that violate the law. VOTER INTENT is the definition we must always follow.

To do this right, we must be prepared to:

- Reject any ballot that may contain "identifying marks."
- Recognize intent when mark is atypical and/or far from target.
- Accurately identify when a vote has been cancelled.

Status

- All 6,737 ballots now online at:
  http://perfect.cse.lehigh.edu/BallotTestData_MNChallengedBallots.html

- Ground truth collected from 8 test subjects, 980 ballot sides.

See DAE demo later today.
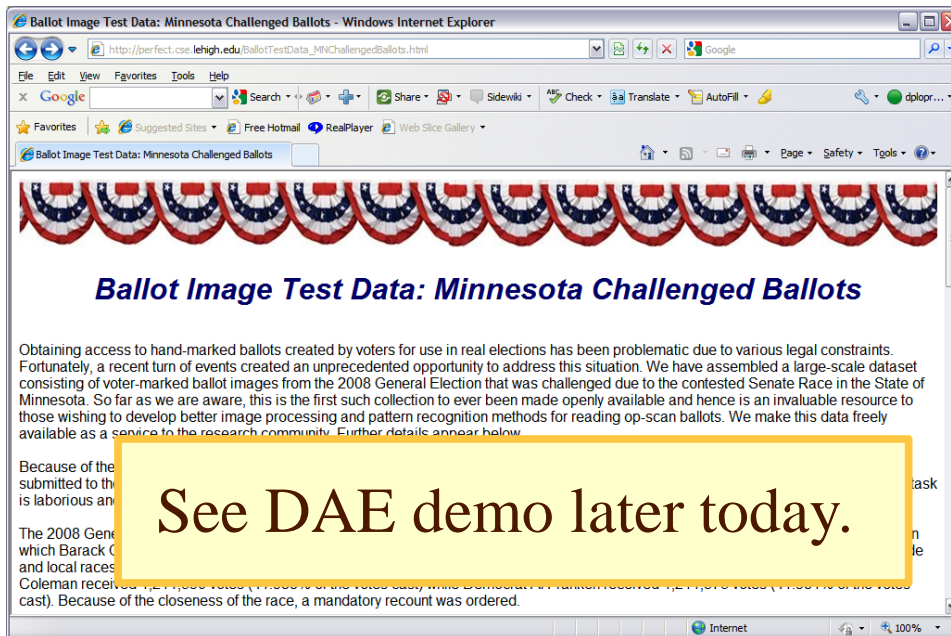
Coming soon:

- Add to TC-11 website.

- Ground-truthing continues – to be made available online.