Instructions for Ground-Truthing OpScan Ballot Images  
PERFECT Project  
http://perfect.cse.lehigh.edu/  

Daniel Lopresti  
Department of Computer Science & Engineering  
Lehigh University  

January 18, 2010

1. Overview

After the contentious 2000 U.S. Presidential election, there was a push for widespread adoption of electronic voting equipment. Computer security experts and concerned citizens, however, have begun raising serious questions about the reliability and trustworthiness of such systems when collecting, storing, and tabulating votes. Direct Record Electronic (DRE) voting, once seen as a straightforward albeit expensive solution, is increasingly viewed as an unacceptable compromise. In its place, there is a growing call for voting systems that employ some form of paper artifact to provide a verifiable physical record of a voter’s choices. Often, this takes the form of a hand- or machine-marked paper ballot which is processed by an optical scanning (opscan) system and then safely secured in the event a recount becomes necessary.

To support research on opscan voting, our goal is to create ground-truth files for a large collection of real ballots from the recount that took place in the 2008 U.S. Senate race in Minnesota. These disputed ballots were scanned and the images were made available to the public on the WWW. In assisting us in this effort, you will be using a software tool we have developed known as perfect.tcl which allows you to record your interpretation of the markings made by voters on their ballots. Your interpretation should be guided by your understanding of the Minnesota state law regarding “voter intent” which is cited later in this guide, along with the information you need to install the software, ground-truth the ballot images, and return the results to us. You will also see screen snapshots of the perfect.tcl tool in action. An example of a fully-annotated ballot image appears at the end of these instructions.

2. Installing the Software

You will probably want to create a special folder on your computer for this work – place all of the files you receive in the same folder. This will include the perfect.tcl application, one or more configuration files, and a set of election specification files (these are basically just lists of ballot images along with some default parameter settings).

To run the perfect.tcl program, you will first need to get and install three free software packages, if you do not already have them on your system. (You will probably need to install all three packages under MS Windows – on Linux systems, you may find that some of these are already present.)

Here is what you will need:

(1) ActiveState Tcl/Tk:  
http://www.activestate.com/Products/activetcl/index.mhtml
(2) The netpbm library:
http://gnuwin32.sourceforge.net/packages/netpbm.htm

(3) The TIFF library:
http://gnuwin32.sourceforge.net/packages/tiff.htm

When using Windows, after you have installed these packages you must do one more thing – you need to update the Windows environment variable to point to the library routines. You do this by selecting “System” in the Windows Control Panel. Then click on the “Advanced” tab. Then click the button toward the bottom labeled “Environment Variables”. Under “User variables”, click “Edit.” You need to add the following path to whatever is already present:
;C:\Program Files\GnuWin32\bin;

![Setting environment variables under MS Windows.](image)

This assumes that netpbm and libtif installed their binaries in the “Program Files/GnuWin32/bin” directory on your C: drive, which should be the default. On 64-bit and/or non-Intel architectures, this path may be “Program Files (x86)/GnuWin32/bin.” (Note that the semicolon character is used as a separator above.) Be sure to click “OK” to save your changes through the various dialog boxes.

3. Running the Application

Double-click on the perfect.tcl application and it should start up. If requested, enter your user ID (this
can be any unique ID that identifies you – please be sure to use the same ID from session to session). Under the “Election” menu, selection “Open...” and navigate to select one of the election specification files you were given. Once an election is open, you will be presented with two windows: the image of the first ballot in the set as well as an “Annotations” window. You will interact with the ballot image by drawing enclosing rectangles and other objects directly on the screen. The “Annotations” window, shown below, encodes information for each race and each candidate on the ballot. Clicking “Next Ballot” and “Prev Ballot” takes you forward and backward through the ballots in the set.

Screen snapshot of the “Annotations” window showing candidates in the Presidential race.

4. Ground-Truthing OpScan Ballot Images

Using the “Annotations” window, record your interpretation for each target oval on the ballot. Most of the ovals will probably be unmarked (“No Vote”). In the case of marked target ovals, after you select an interpretation by clicking on the associated radio button, you will be given a pencil tool to draw a rectangle around the the mark in question on the ballot. Your rectangle should completely enclose the mark and it should include no other black pixels on the ballot, other than the original mark target which is the oval that was pre-printed on the ballot. It is not necessary for the rectangle to be “tight” around the mark, however. You may find you need to zoom in to draw an accurate rectangle.

The rectangle in the middle of the screen snapshot below is too large and overlaps with a pre-printed ruling line, while the rectangle on the bottom does not include the entire mark made by the voter. The top rectangle is drawn correctly.
A correctly marked target (top) and two badly marked targets (middle and bottom).

Sometimes it is not possible to follow these guidelines because the voter’s marking extends beyond the target oval and the white space around it into pre-printed text. In such cases, draw your rectangle to completely enclose the voter's mark, no matter what else might be included. This is illustrated below.

Keep in mind that the rectangles you draw are selectable, movable, and resizable. You can also copy and paste markup. Rectangles that are mistakes can be deleted. If you wish to reassign an existing rectangle to a different interpretation, select the rectangle on the ballot image and then select the correct new interpretation in the “Annotations” window while holding down the “Shift” key the entire time.

In addition to ground-truthing individual target ovals and marks, you should also indicate the “winner” or “winners” (if any) in each race on the ballot. This information may be implicit if there is only one valid mark, but there could in fact be multiple votes and other markings in a single race; hence, the question of who deserves the legal vote cannot always be resolved by examining the annotations for each target oval in isolation. What is required is a high level human interpretation of the voter's intent. To accurately determine this, you must keep in mind the Minnesota guidelines which are listed later.
Next to each candidate's name in the annotation panel you will see a checkbox that says “Legal Vote.” This is the place where you will record the voter's intent. Note that it is quite possible for there to be two valid markings in a single race, but no “Legal Vote” (if the race is an overvote, then all of the votes in that race are invalidated). It is important to keep in mind that some races allow more than one choice. The “Vote For” field, which you will see in the annotation panel, reports the maximum number of candidates who can receive a vote.

A race with a cancelled vote (Coleman) and a valid marking (Franken). Franken is recorded as receiving the “Legal Vote” in this case.

You may find that certain races on the ballot image do not match the entries in the annotation file. This arises when the annotation file was created for a different county. In general, the national and statewide races will be the same across all ballots, but certain local races will differ from county to county. In that case, you will create a simple Microsoft Excel spreadsheet file that represents the ballot you are viewing and import this into the perfect.tcl tool to create the appropriate annotation file. The instructions for this procedure are described in a later section.

Please ground-truth the entire ballot image. In addition to recording votes, look for various random marks that may appear and annotate them using the appropriate rectangle from the “Markup” menu. Textual annotations can be added on the ballot by selecting the “Add Label” command. The following types of markup are provided:

- Valid Vote
- Erased Vote
- Bleed-Through
- Handwriting (voter)
- Handwriting (unknown)
- Label
- Fiducial
- Cancelled Vote
- Stray Mark
- Challenge Stamp
- Handwriting (official)
- Other Mark
- Skew Line
Below is an example of a stray mark:

Example of a stray mark.

When voters realize they have made a mistake, they sometimes try to cancel a vote by writing a large “X” through it, as was shown earlier. Other times, it appears as though they have attempted to erase the marking, as illustrated below. Use the “Cancelled Vote” annotation for each of these cases and, in addition, if you believe an erasure is involved, you may overlay that annotation as well.

Attempting to cancel a vote by erasing the marking.

Judgment must be used when interpreting voter intent, since a marking that appears to be an attempt to cancel a vote on one ballot could in fact be an intended vote on another ballot, as shown below.
A valid vote that gives the appearance of being cancelled.

To the extent it is possible, we would like you to try to classify the handwriting you see on the ballot as belonging to the original voter (added at the time of the election) or to election officials (added after the election). In theory, voters should never write on their ballots. However, since these are challenged ballots, some of the writing you see will indeed belong to the voter. We realize that sometimes it may be impossible to tell who the writing belongs to – in that case, just use plain “Handwriting” annotation. In the case below, we can be fairly certain that the handwriting belongs to the voter, and that he/she was attempting to indicate a desire to cancel the partial marking of the oval corresponding to Al Franken. (It can be argued that this marking could be used to identify the voter and his/her votes after the election, which would invalidate the ballot in question according to Minnesota state statutes.)

Example of voter writing on the ballot.

Bleed-through can occur when the voter makes heavy marks on one side of the ballot that show up on
the other side. Note that you may need to examine both the ballot front and back to make a
determination. The `perfect.tcl` tool allows you to select the ballot side under the “Display” menu.

Example showing bleed-through from the back of the ballot to the front.

You can also add free-form text using a box at the bottom of the “Annotations” window. For example, if you see writing on the page, or other unusual situations, you can enter a note into this text box.

Although each ballot has two sides, for the majority of the work you do you will be concentrating on
the front of the ballot. This is easily identified as the side that contains the U.S. Presidential race, as well as the challenged U.S. Senate race. However, there is one piece of important information that you will normally find on the back of each ballot, and this is the Challenge Stamp, as shown below. In addition to annotating the location of this stamp, please also enter the name of the challenger and the reason for the challenge in the free-form text box at the bottom of the “Annotations” window.

Annotating the challenge stamp on the back of the ballot.
(When switching to view the back of the ballot, **perfect.tcl** may complain that it cannot find an annotation file and is using the default — this is not a problem, just click “OK” and proceed.)

Before finishing with each ballot image, be sure to add a “Skew Line.” Overlay this line with a long printed line on the ballot image that you think should be horizontal. This allows us to estimate the angle at which the ballot was scanned.

Skew line overlayed on a long horizontal ruling line on the ballot image.

Paper forms designed for opscan processing often contain fiducial markings. These are landmarks placed at fixed positions on the page to allow the scanned image to be registered with a “clean” copy of the form. The Minnesota ballots we are using contain fiducial markings in each of the four corners, although the marks are small and are not present in all of the images. These marks look like small circles containing “cross-hairs,” although sometimes they are hard to see because of the quality of the scan. When possible, please annotate the four fiducial markings on the ballot as well, aligning the center of the cross-hairs as best you can.

Annotating fiducial markings.

Because the **perfect.tcl** tool allows you to move backwards and forwards through the ballot set, you do not need to complete the ballots in any specific order, and you can go back and fix mistakes if needed. The tool also provides the ability to list all of the ballots in a set and jump to one in the middle of the list without having to click “Next” and “Prev” a large number of times. To access this feature, go to “List Ballots...” in the “Election” menu.
Please spend some time experimenting with `perfect.tcl` before you use it to start ground-truthing real ballots. To do so, you can use the practice set that we provide.

5. Interpreting Votes

For this data collection activity, we are asking you to perform two different but related tasks:

1. We want you to interpret each target oval and its associated markings individually. Hence, it is possible for a given race to have two valid marks, even though that might invalidate the ballot from the standpoint of that particular race. When deciding whether a marking is an intentional vote, a stray mark, or an attempt to cancel or erase a mistake, you may consider the other marks made by the voter on the ballot. Remember, however, to interpret each target oval individually.

2. We also want you to decide who – if anyone – receives the legal vote(s) in each race. This is the normal way that a manual recount is conducted. If a race is overvoted (the voter selects more candidates than the maximum specified), then there is no legal vote in that race. Undervoting (selecting fewer candidates than the maximum specified) is always permissible.

The ballots you will be ground-truthing come from the challenged U.S. Senate race in this 2008 election. Hence, in some cases (but not all), markings in that particular race may be atypical.

![Image of a ballot for School Board Member Independent School District No. 1 (Aitkin)](image)

A race with three legal votes.

You should review the official Minnesota statutes for interpreting voter intent which are listed below, keeping in mind once again that some of these are intended to determine the winner (if any) of a particular race, and not whether a marking in isolation appears to be valid.

```
2008 MINNESOTA STATUTES
(from https://www.revisor.leg.state.mn.us/statutes/?id=204C.22)

204C.22 DETERMINING VOTER'S INTENT.
```
Subdivision 1. **Ballot valid if intent determinable.**
A ballot shall not be rejected for a technical error that does not make it impossible to determine the voter's intent. In determining intent the principles contained in this section apply.

Subd. 2. **From face of ballot only.**
Intent shall be ascertained only from the face of the ballot.

Subd. 3. **Votes for too many candidates.**
If a voter places a mark (X) beside the names of more candidates for an office than are to be elected or nominated, the ballot is defective with respect only to that office. No vote shall be counted for any candidate for that office, but the rest of the ballot shall be counted if possible. At a primary, if a voter has not indicated a party preference and places a mark (X) beside the names of candidates of more than one party on the partisan ballot, the ballot is totally defective and no votes on it shall be counted. If a voter has indicated a party preference at a primary, only votes cast for candidates of that party shall be counted.

Subd. 3a. **Votes yes and no.**
If a voter votes both yes and no on a question, no vote may be counted for that question, but the rest of the ballot must be counted if possible.

Subd. 4. **Name written in proper place.**
If a voter has written the name of an individual in the proper place on a general or special election ballot a vote shall be counted for that individual whether or not the voter makes a mark (X) in the square opposite the blank.

Subd. 4a. **Write-in vote for candidate team.**
A write-in vote cast for a candidate for governor without a write-in vote for a candidate for lieutenant governor must be counted as a vote for the candidate team including the lieutenant governor candidate selected by that candidate for governor.

Subd. 5. **Name written on primary ballot.**
If a voter has written the name of an individual on a primary or special primary ballot, a vote shall not be counted for that office.

Subd. 6. **Mark out of place.**
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.

Subd. 7. **All written names or marks counted up to limit.**
If a number of individuals are to be elected to the same office, the election judges shall count all names written in and all printed names with (X) marks in squares opposite them, not exceeding the whole number to be elected. When fewer names than the number to be elected are marked with an (X) or written in, only the marked or written in names shall be counted. When more names than the number to be elected are marked or written in, the ballot is defective with respect to that office and no vote shall be counted for that office.

Subd. 8. **Misspelling; abbreviations.**
Misspelling or abbreviations of the names of write-in candidates shall be disregarded if the individual
Subd. 9. **Votes for only some offices or questions determined.**
If the voter's choice for only some of the offices or questions can be determined from a ballot, the ballot shall be counted for those offices or questions only.

Subd. 10. **Different marks.**
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 as provided in subdivision 13.

Subd. 11. **Attempted erasures.**
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. If an attempt has been made to obliterate a write-in name a vote shall be counted for the remaining write-in name or marked candidate.

Subd. 12. **Soil; defacement.**
A ballot shall not be rejected merely because it is slightly soiled or defaced.

Subd. 13. **Identifying ballot.**
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.

Subd. 14. **No votes for certain offices.**
If the number of candidates for an office is equal to the number of individuals to be elected to that office, and the voter has not marked any name, no vote shall be counted for any candidate for that office.

Subd. 15. **Blank ballot for one or more offices valid.**
If no name or response to a question is marked and no name is written in, the ballot is blank with respect to that office or question. A ballot that is blank with respect to one or more offices or questions is not defective.

6. Creating New Annotation Files

Across a state, certain races will be common to all ballots (e.g., the Presidential race), whereas other races will be different from county to county or town to town. This means, unfortunately, that a large number of base annotation files are needed to encode the election in Minnesota, one for each possible ballot design. It would be infeasible to prepare all of these in advance. Instead, the perfect.tcl tool allows you to create new annotation files as you find they are needed.

To do this, you will first create (or edit) a simple spreadsheet file representing the ballot using Microsoft Excel (the freely-available OpenOffice suite will also work). The basic format for the file is shown below. In the first column of the spreadsheet, you will list the names of the races found on the
ballot. In the second column of the same row, you will enter an integer that represents the maximum number of legal votes for that race. Below the entry for a given race, in the second column, you will list each of the candidates, one per row. It is very important to use the same spelling as appears on the ballot. Also use the same conventions regarding capitalization. For example, if a word appears in all-caps on the ballot image, then enter it this same way in the spreadsheet. Generally, the ballot image will contain additional textual information beyond the names of the races and the candidates. Do not enter this additional information in the spreadsheet.

Once you have created the spreadsheet file, save it in the file format known as “CSV” (comma separated values). After you have done this, select the “Convert .csv to .note File...” command from the “Perfect” menu in the perfect.tcl application. Browse to the CSV file you saved earlier and select it. The perfect.tcl tool will then convert this file into an annotation file that can be loaded into the application. To load this new annotation file, choose the “Set .note File...” command from the “Perfect” menu and browse to the associated file. The new .note file will be loaded and its entries will appear in the “Annotations” window for you to use in ground-truthing the ballot in question.

7. Other Hints and Advice

We strongly encourage you to make periodic backups of your ballot files. The perfect.tcl application is still a prototype and you may encounter programming errors when using it. If you see an error message while ground-truthing a ballot, the safest course of action is to quit the application immediately, make a backup of the files in question in a safe place, and then restart perfect.tcl to try to pick up where you left off. If the error still occurs, inform us by sending us a description of the error as well as copies of the files you were using at the time. You can continue to ground-truth other ballot images while we attempt to identify the cause of the error.

Screen snapshot of a Microsoft Excel file for inputting a ballot specification.
In addition to recording your interpretation of the ballot markings and the votes in each race, the “Annotations” window also asks you to specify the marking style(s) for the voter. This can be “Filled Oval,” “Partially Filled Oval,” “Ex Mark (X),” “Check Mark,” or “Other.” Select any or all of these options that apply.

The perfect.tcl tool is designed to log the actions you take when ground-truthing a ballot image. This data will be analyzed as part of our studies. Please be sure the “Log Events” option is selected under the “Perfect” menu before you start to ground-truth.

8. Returning Your Results

As you work through each election specification file, perfect.tcl will create an associated folder in your directory. When you have finished ground-truthing, send me these folders via email. Also remember to keep separate notes about bugs you find in the software, unusual cases you see in the ballot collection, and thoughts you have on improving the ground-truthing process.
9. Example of a Fully Annotated Ballot

OFFICIAL BALLOT

STATE GENERAL ELECTION BALLOT

INSTRUCTIONS TO VOTERS

FEDERAL OFFICES

STATE OFFICES

COUNTY OFFICES

COUNTRY OFFICES

CONSTITUTIONAL AMENDMENT

CLEAN WATER, WILDLIFE, CULTURAL HERITAGE AND NATURAL AREAS

COUNTY OFFICES

SCHOOL DISTRICT OFFICES

VOTE FRONT AND BACK OF BALLOT