Evaluation of an Automated Pavement Distress Identification and Quantification Application

Jerome Daleiden, Nima Kargah-Ostadi (Fugro)

Abdenour Nazef (Florida DOT)

Road Profile User's Group (RPUG) Conference

November 3, 2016

San Diego, CA

this presentation will address:

- background on distress survey methods
- framework for evaluating existing methodologies
- evaluation results for FDOT rigid pavement surveys
- context-sensitive gap analysis
- recommendations

Pavement Distress Survey Methods

- **Manual** (windshield or walking)
  - Safety concerns
  - Agreement issues among raters
- **Semi-Automated** (manual review of images)
  - No safety concerns
  - Agreement issues among raters
- **Automated** (detection and classification software)
  - Set reasonable accuracy goals and QA procedure
  - Still requires post-survey QC

Existing FDOT Rigid Pavement Survey Protocol

- Transverse Cracking (count), Light-Moderate-Severe
- Longitudinal Cracking (count), Light-Moderate-Severe
- Spalling (linear feet), Moderate-Severe
- Corner Cracking (count), Light-Moderate-Severe
- Patching (sq. yards), Fair- Poor
- Shattered Slabs (count), Moderate-Severe
- Surface Deterioration (sq. feet), Moderate-Severe
- Pumping (percentage range: Code 1 to 4), Light-Moderate-Severe
- Joint Condition, partially sealed, not sealed
- Multiple Cracked, Slabs count

Distress Identification Workshop

- Increase consistency among raters
- Notes (clarifications) on existing protocol
- 80% agreement among raters (3) for total transverse cracking
- 63% agreement for total longitudinal cracking
- 65% agreement for spalling
- More agreement in total amount of each distress type than on distress severity levels

Evaluation Framework

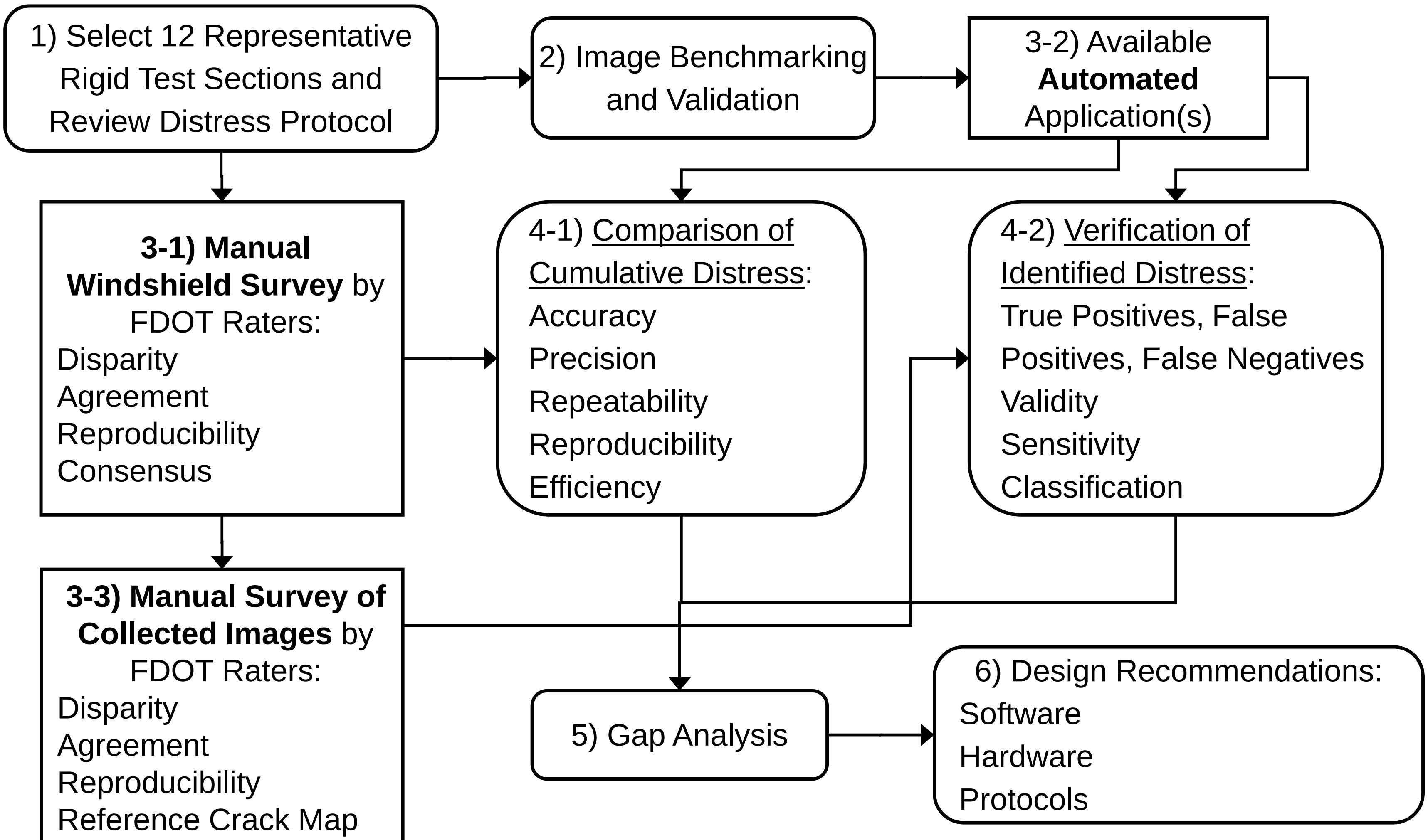


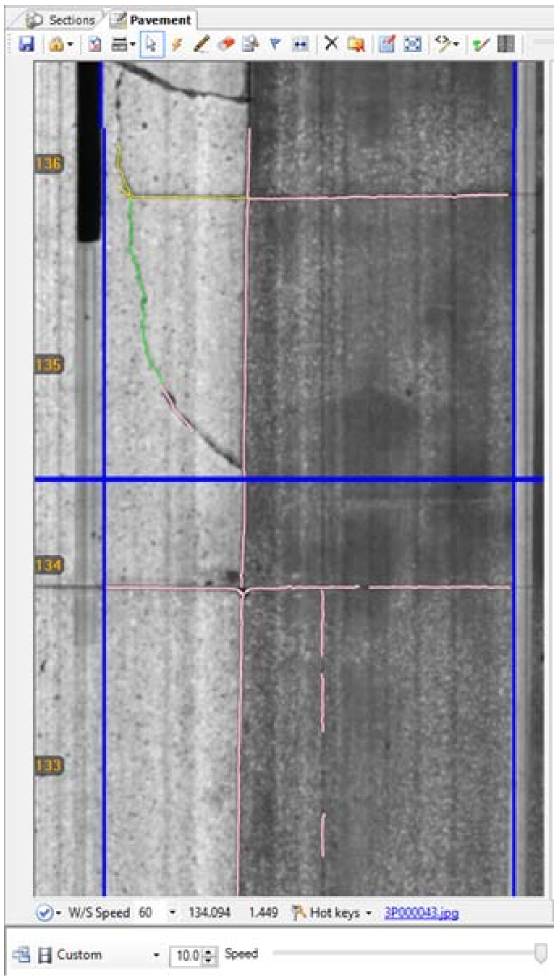
Image Quality Validation

- **Image Properties**: resolution, exposure, dynamic range, white balance
- **Image Issues**: alignment of control lines, image streaks
- **Image Feature Capturing**: optical distortion, signal-to-noise ratio
- **Hardware**: distance measuring accuracy, latitude-longitude accuracy, platform stability
- **Environmental Effects**: lighting conditions, temperature, humidity, wind, vehicle speed
- **AASHTO PP68 →** depends on crack detection software

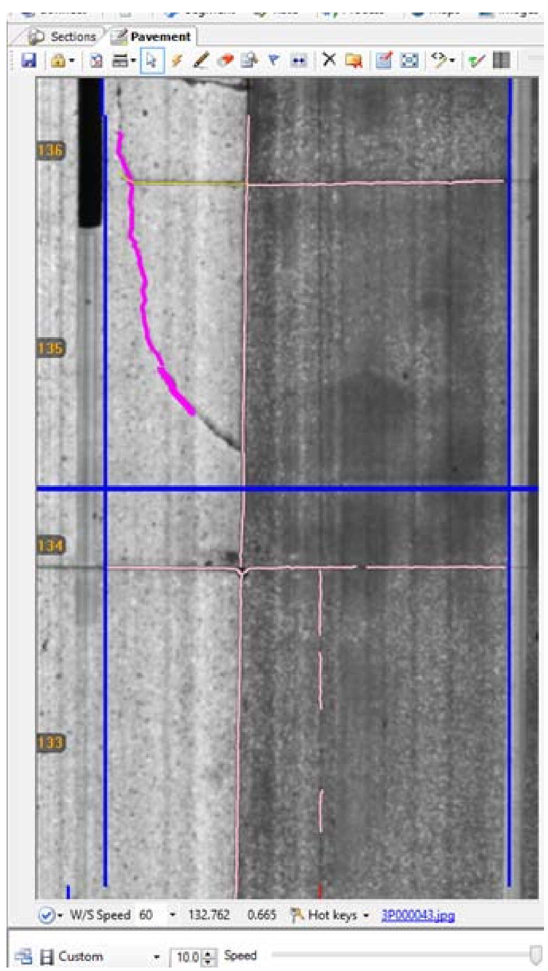
Optimum Software Settings

1) Image Pre-Processing: Applying Filters
2) Detection (find optimum parameters)
3) Classification (and grouping)
4) Rating Distress Type and Severity

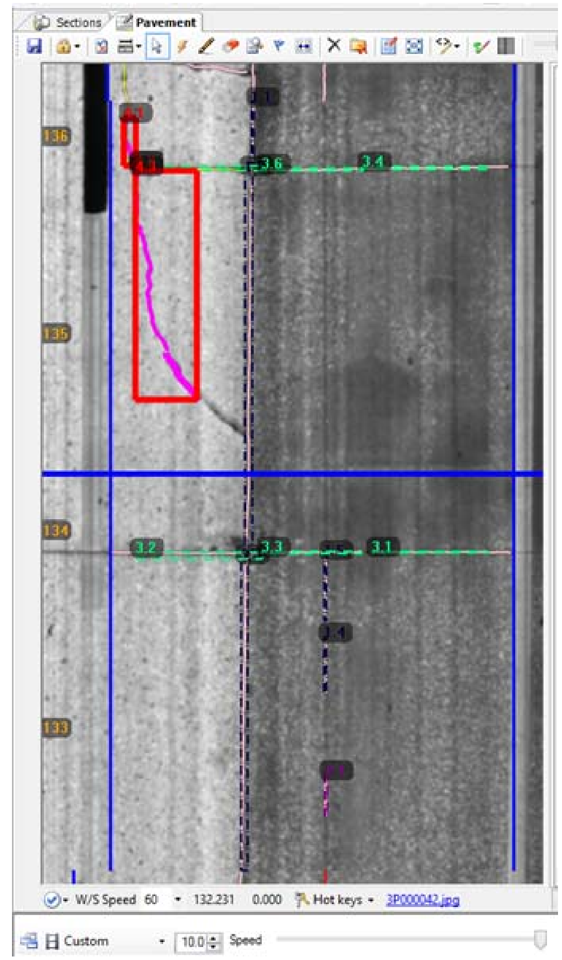
Automated Survey Steps



Detection



Classification

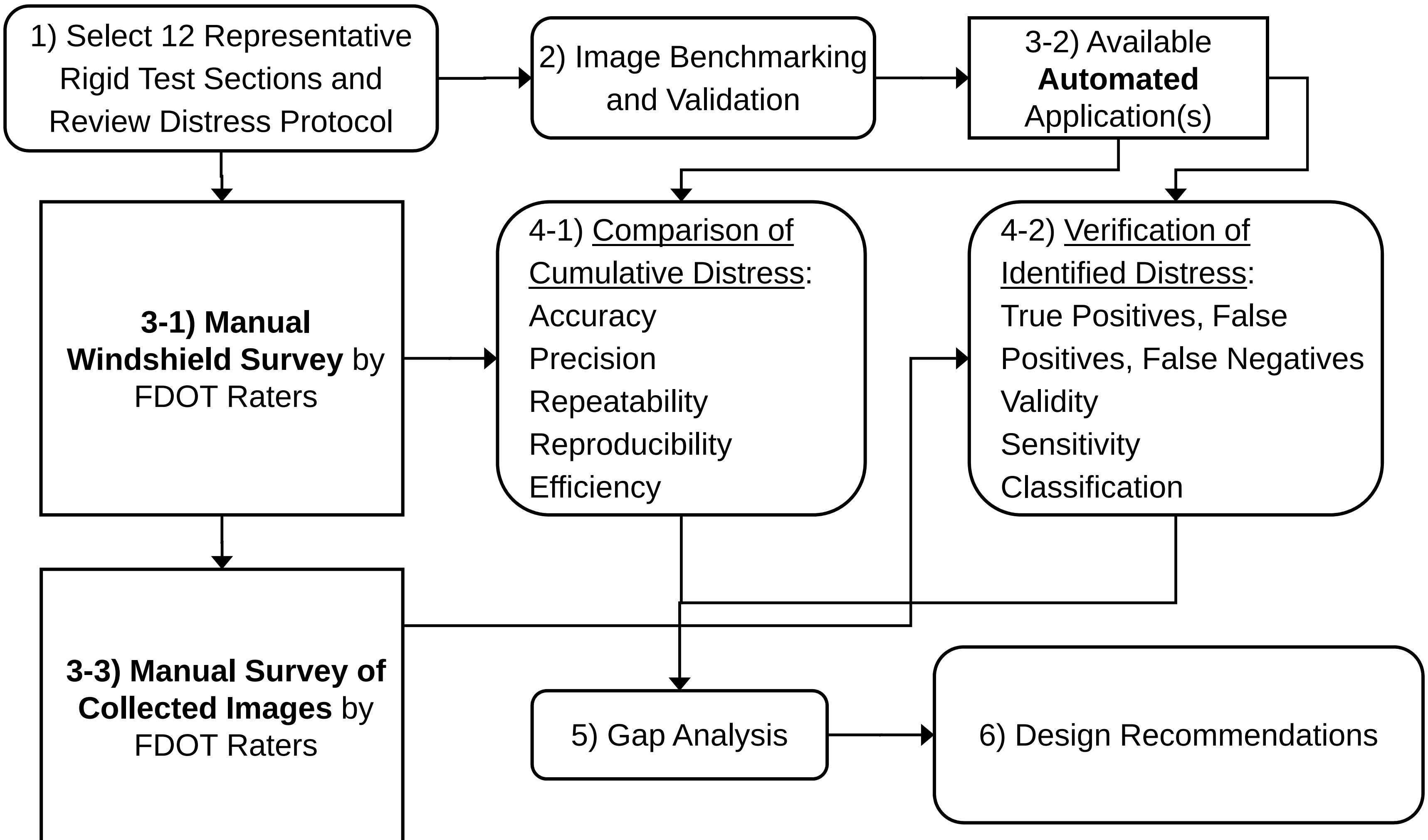


Rating

Automated Rating

- **Transverse Joints**
- **Longitudinal Joints**
- Issues with skewed joints and transition areas
- **Transverse Cracks**
- **Longitudinal Cracks**
- Cracks per slab
- Grouping and counting cracks

Evaluation Framework



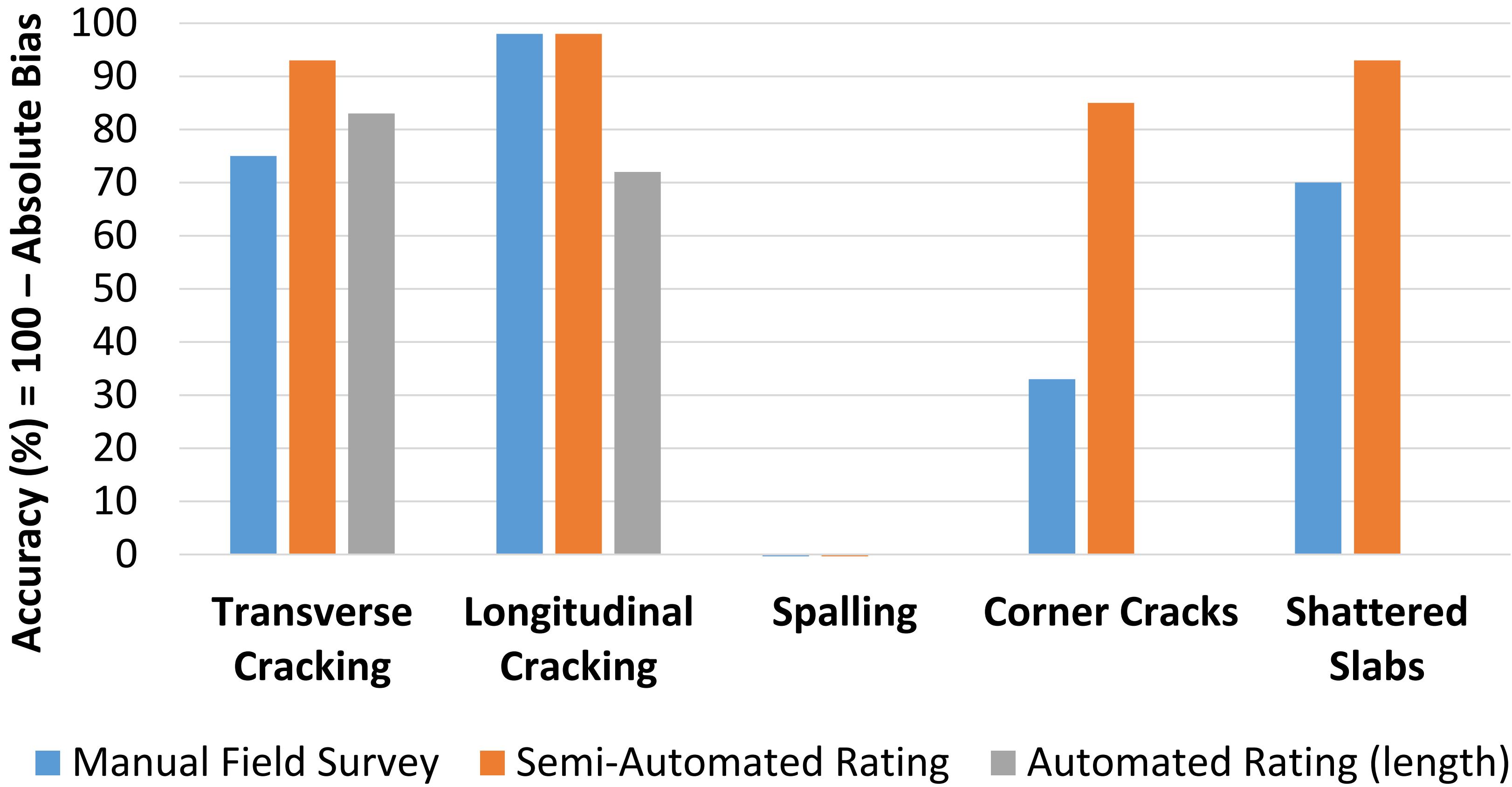
Success Metrics

- **Effectiveness**: Accuracy
- **Efficiency**: Speed
- **Reliability**
  - **Precision**: variation among different sections
  - **Reproducibility**: variation among raters
  - **Repeatability**: variation among runs

Reference values

Accuracy

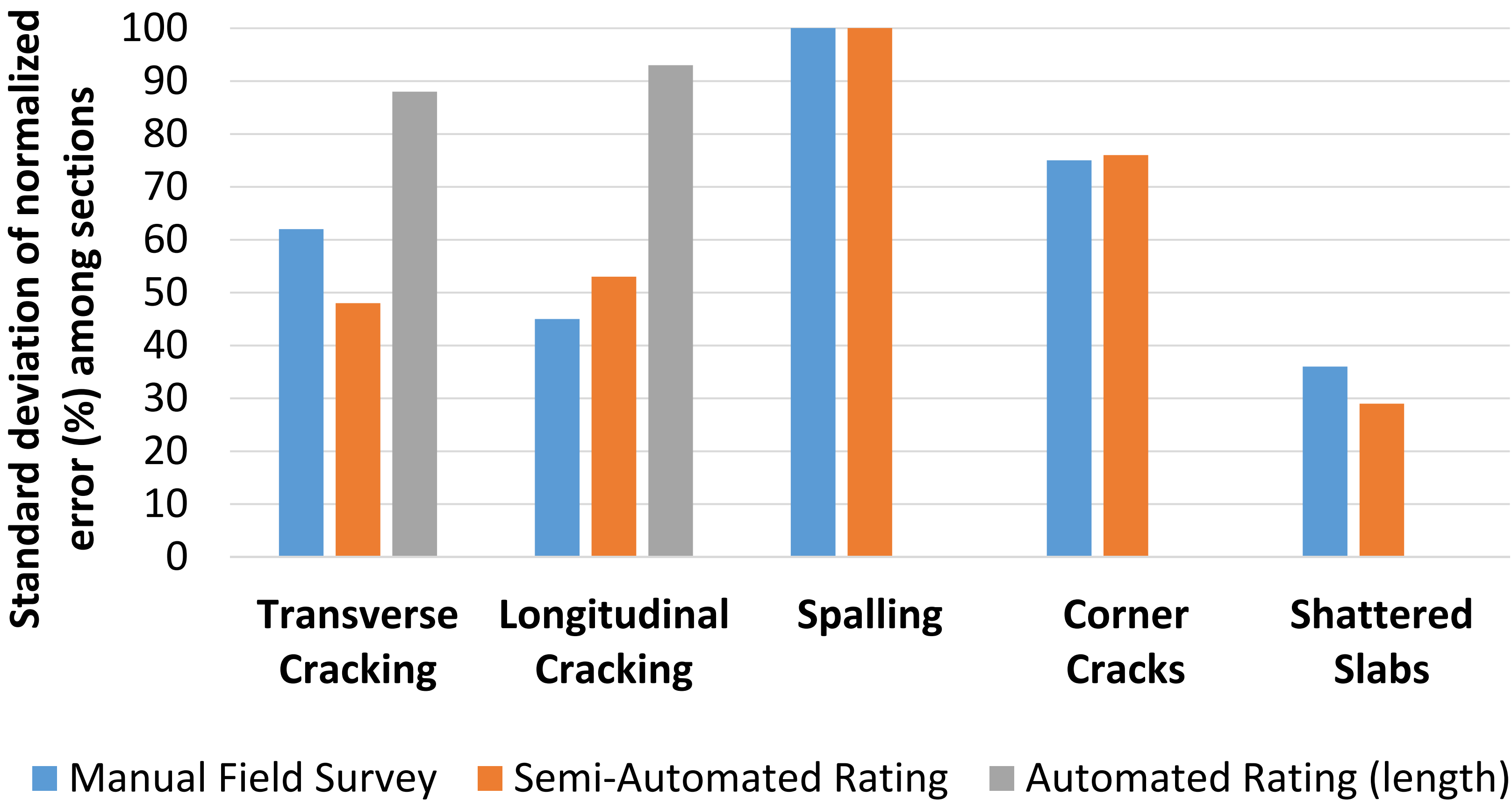
Bias (average normalized error) compared to a reference rating



spalling bias was over 100%

Precision

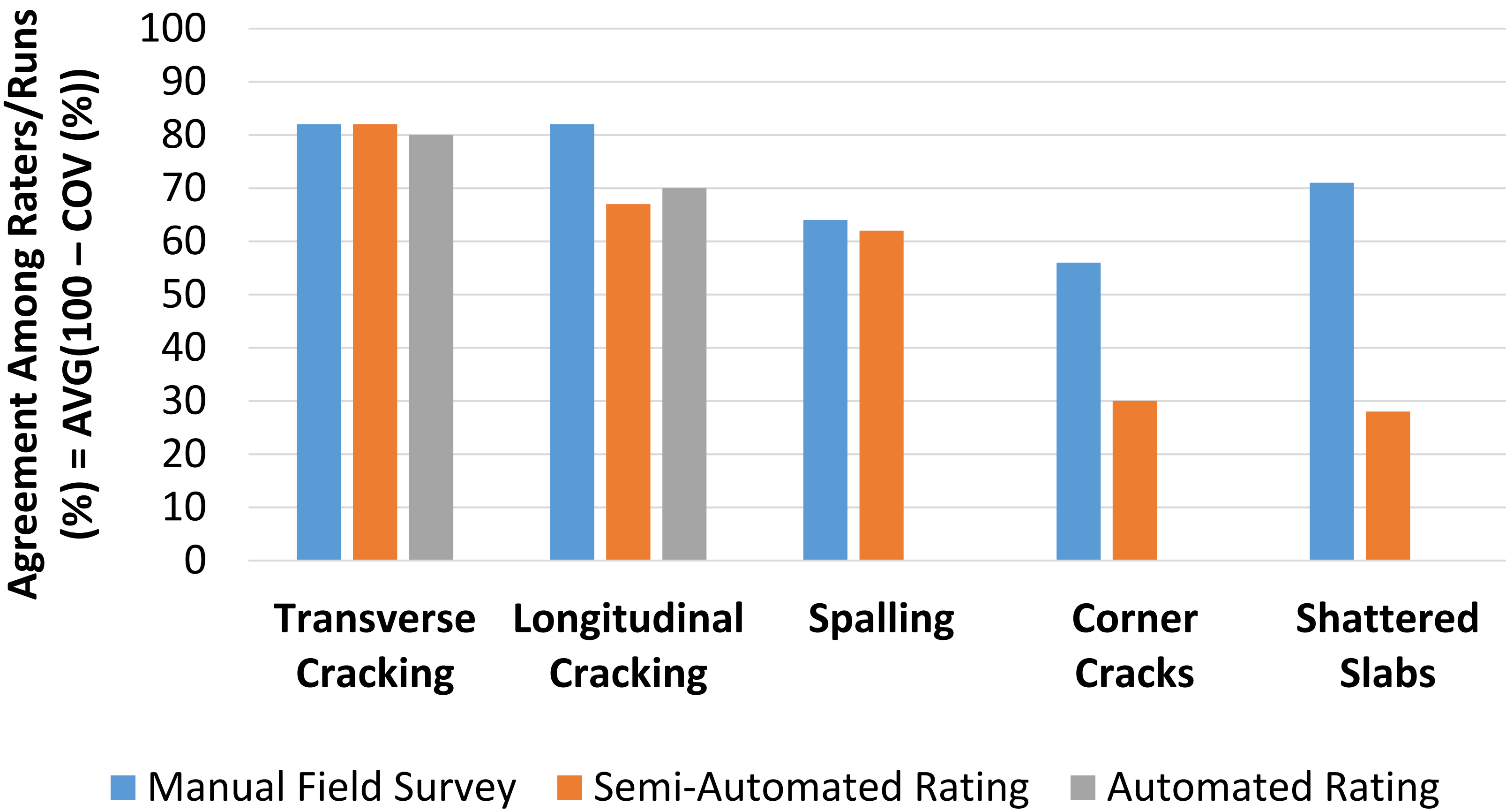
Variation of error among sections



spalling variation was above 100%

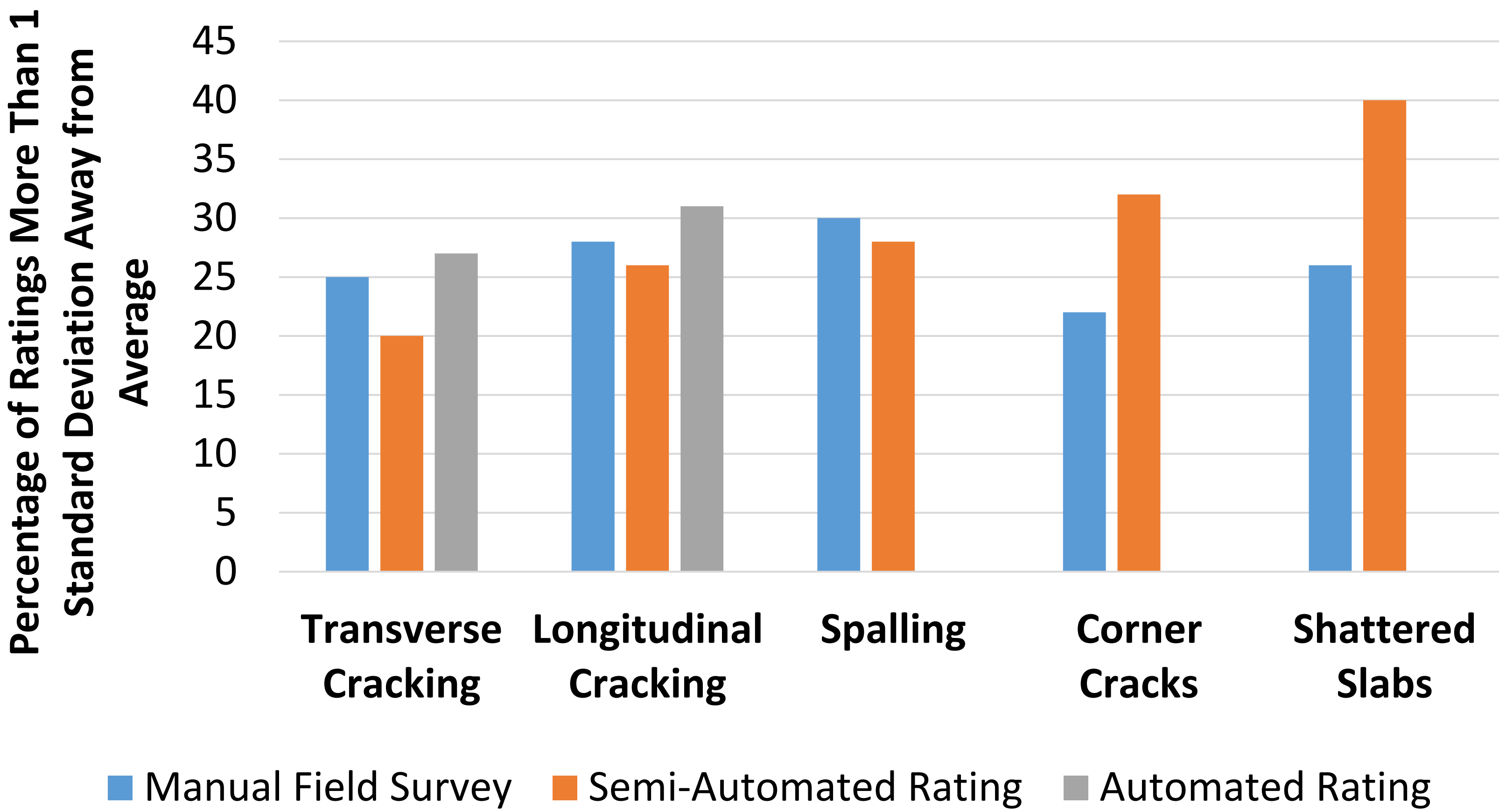
Reliability

Variation of error among raters/runs

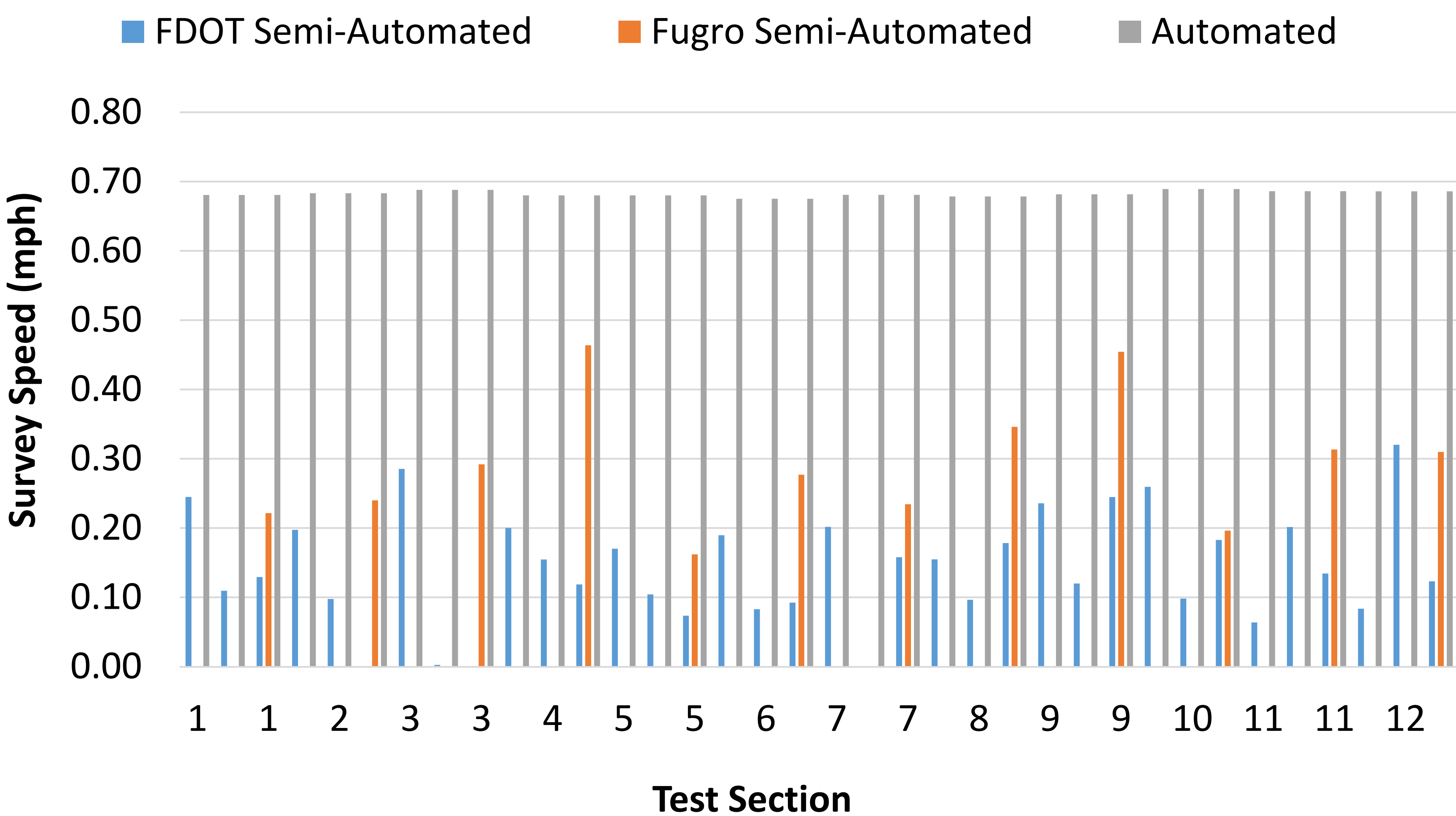


“Outlier” Ratings

A measure of rating reproducibility/repeatability



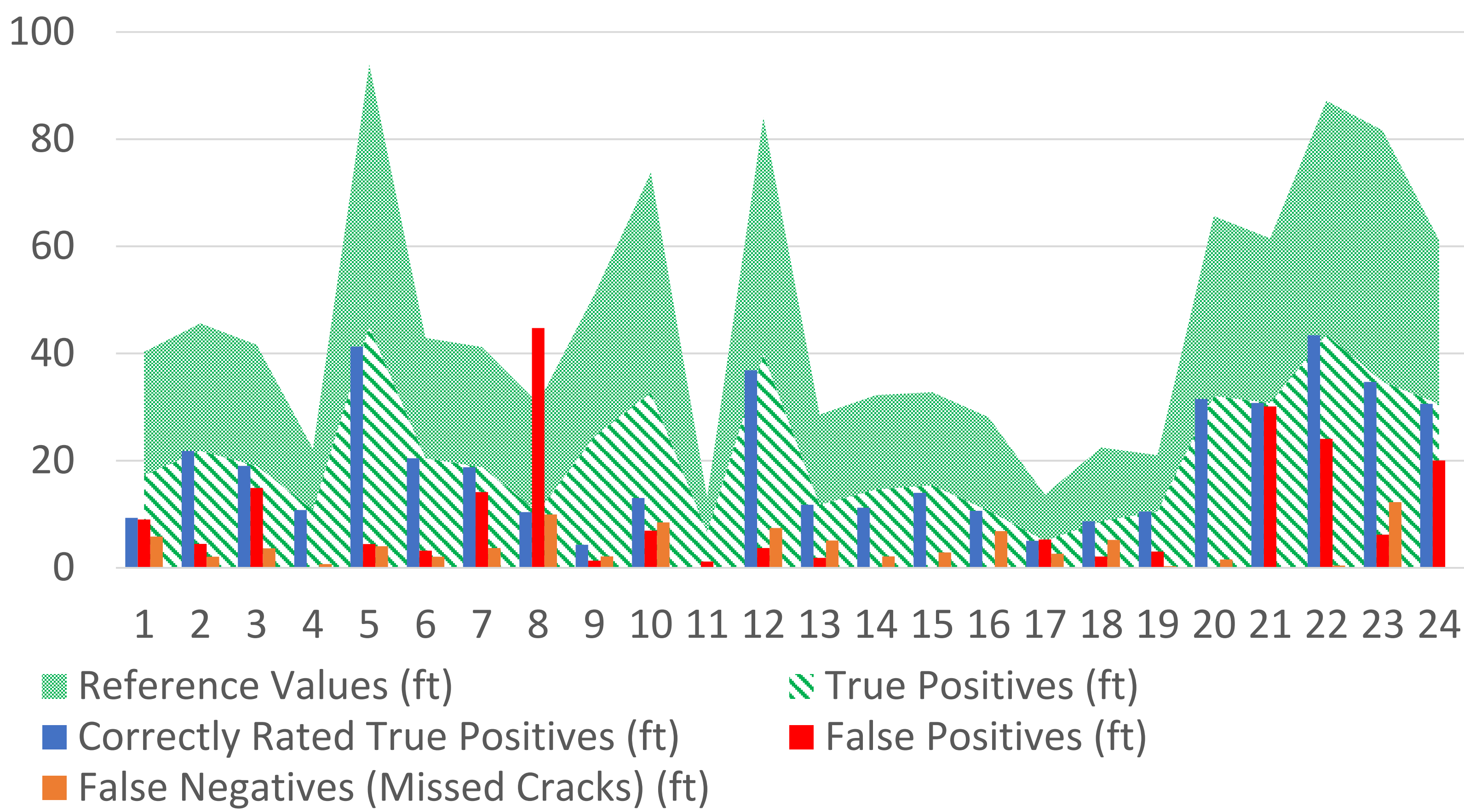
Efficiency



FDOT Semi-Automated
Fugro Semi-Automated
Automated
Survey Speed (mph)
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1 1 2 3 3 4 5 5 6 7 7 8 9 9 10 11 11 12
Test Section

Verification of Detected Distress

Results on 24 random image frames



Verification of Detected Distress

Average results on 24 random image frames

Statistic	Length (ft)					Count	
	Ground Truth	True Positive		False Positive	False Negative	Reference	Rating Result
		Correctly Rated	Incorrectly Rated				
AVG	25.06	18.68	2.76	8.35	3.70	2.42	6.42
STD	12.49	12.38	5.72	11.25	3.34	1.41	3.12
MIN	6.58	0.00	0.00	0.00	0.00	1.00	2.00
MAX	48.97	43.381	20.202	44.74	12.25	7.00	14.00

Statistic	Distress Validity (or Precision)	Distress Sensitivity (or Recall)	Distress Classification Performance
AVG	78%	84%	86%
STD	21%	14%	28%
MIN	19%	51%	0%
MAX	100%	100%	100%

Key Observations

- semi-automated results show higher accuracy and precision, but lower efficiency and less agreement among raters compared to the manual surveys
- automated survey was more successful in identifying transverse cracks than longitudinal cracks
- 78% of the detected distress was actually present in the reference survey
- Automated method detected 84% of reference survey distresses
- 86% of the detected distress is correctly classified by the automated algorithm

Gap Analysis

Category	Gap	Recommended Solution
Human Random Errors	High variation of rating results among test sections	N/A
Human Systematic Errors	High bias (average error) and high variation of rating results among multiple raters	Review and/or revise distress protocols
Software Systematic Errors	High bias in longitudinal cracking amount (high number of false positives)	Joint detection plugin and plugin for separating stripes
	High variation of error among multiple test sections	
	Issue with crack counts	Improve crack grouping
	Not rating corner cracks	Corner crack plugin
	Not rating shattered slabs	Shattered slab plugin
	Issue with crack width determination and severity rating	Do not use filters moving forward so that crack width can be measured
Hardware Limitations	distresses such as spalling or patching cannot be detected without 3D data	Evaluate 3D data

Design Recommendations

1. Develop plugin for transverse and longitudinal joint detection
2. Develop plugin for lane marking detection
3. Plugin to improve crack grouping and count per slab
4. Plugin to classify and rate corner cracks
5. Plugin to classify and rate shattered slabs

Summary

- Framework for evaluation of different distress survey methods
  - On cumulative amount of distress
    - Accuracy (bias)
    - Precision (variation among sections)
    - Reproducibility/repeatability (variation among raters/runs)
    - Efficiency (speed)
  - Distress-by-distress verification
    - Validity = true positives / (true positives + false positives)
    - Sensitivity = true positive / (true positive +false negative)
    - Classification performance = correctly classified / true positives
- Gap Analysis
- Design Recommendations