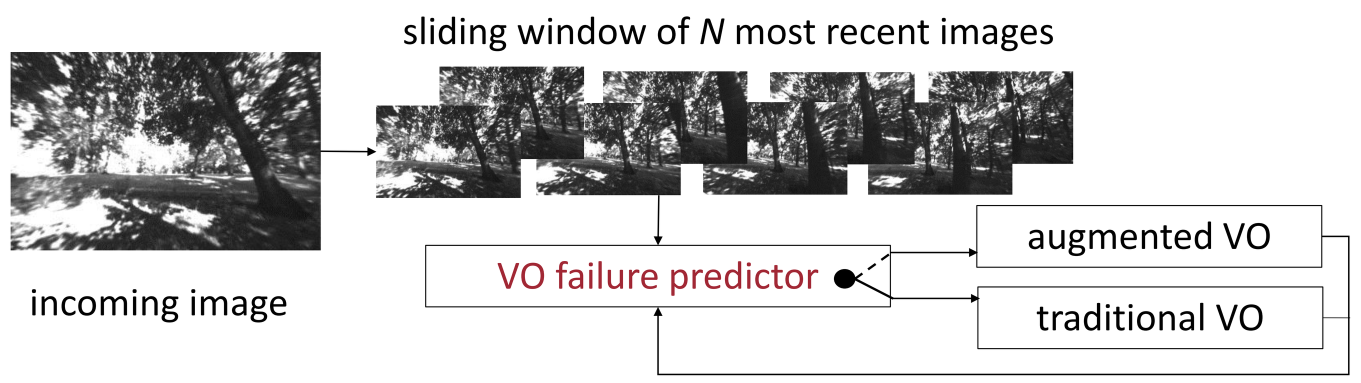
Predicting Feature-Based Visual Odometry Failure using Luminance Saliency

Michael Lee | John Yao | Nathan Michael

Motivation and Objective

- Count, distribution, and quality of features impact the performance of feature-based visual odometry (VO) techniques
- These measures are directly correlated to relevant information available in incoming images

Objective: Predict imminent failures in feature-based VO techniques by detecting degradation in relevant information in *N* most recent images



Framework for introducing robustness to VO

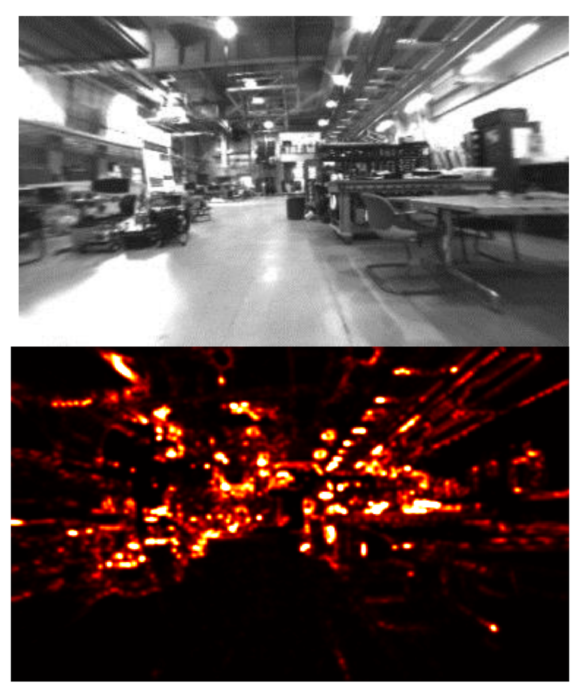
Background

Saliency: Conspicuity in a visual field arising from center-surround contrast [1]

Relevance of saliency: Feature-based VO infers ego-motion from movement of conspicuous points (features)

Selection of saliency type: Luminance

- Luminance was most predictive measure of VO performance among color and depth channels [2]
- Single channel chosen under size, weight, and power constraints



FRC highbay and corresponding saliency map

Approach Overview

- Collect representative datasets with KLT feature-based stereo visual odometry (SVO)
- Find and analyze common failure modalities and corresponding saliency maps
- Create a suite of visual measures to detect common failure modalities
- Train a regression offline, mapping from visual measures to occurrences of imminent failure *M* frames away
- Test on non-overlapping dataset

Dataset Collection and Analysis

Hardware:

- 2 mvBlueFox MLC200wG cameras with baseline of 15.85 cm
- 376 x 240 pixel images published at 10 Hz

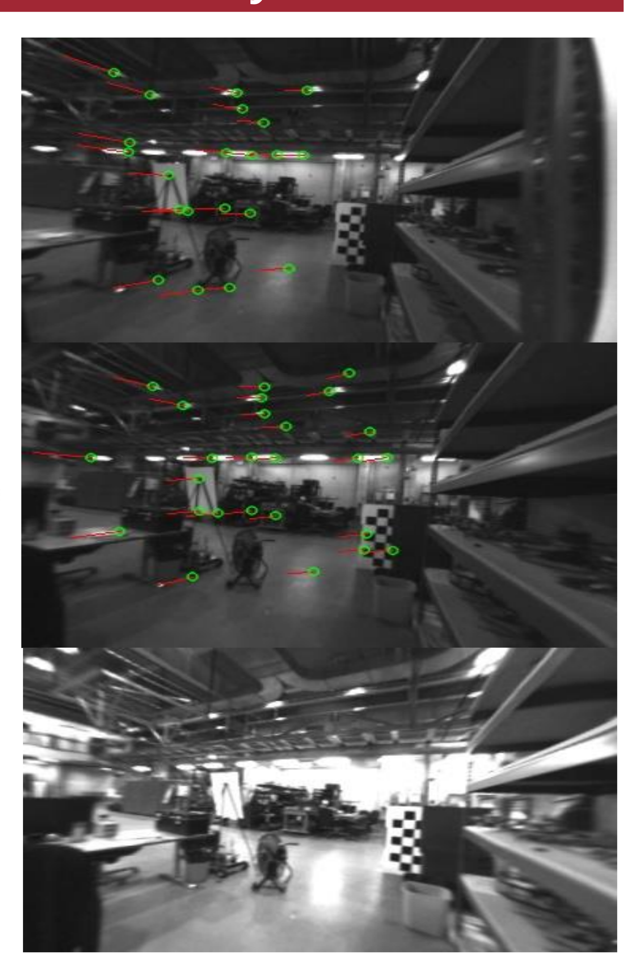
Environments of interest:

- Indoor: office, classroom, hallway, basement, garage
- Outdoor: CMU campus, forest, field, parking lot, road (sunny, overcast weather)
- 77,000 images collected
- 120 recorded instances of failure

Failure: Occurs when zero features are tracked between any two consecutive frames

Observed failure modalities: All recorded failures correlated with one or more of the following modalities

A. Lack of sufficient contrast (36%)
B. Sudden exposure change (47%)
C. Rapid movement (17%)



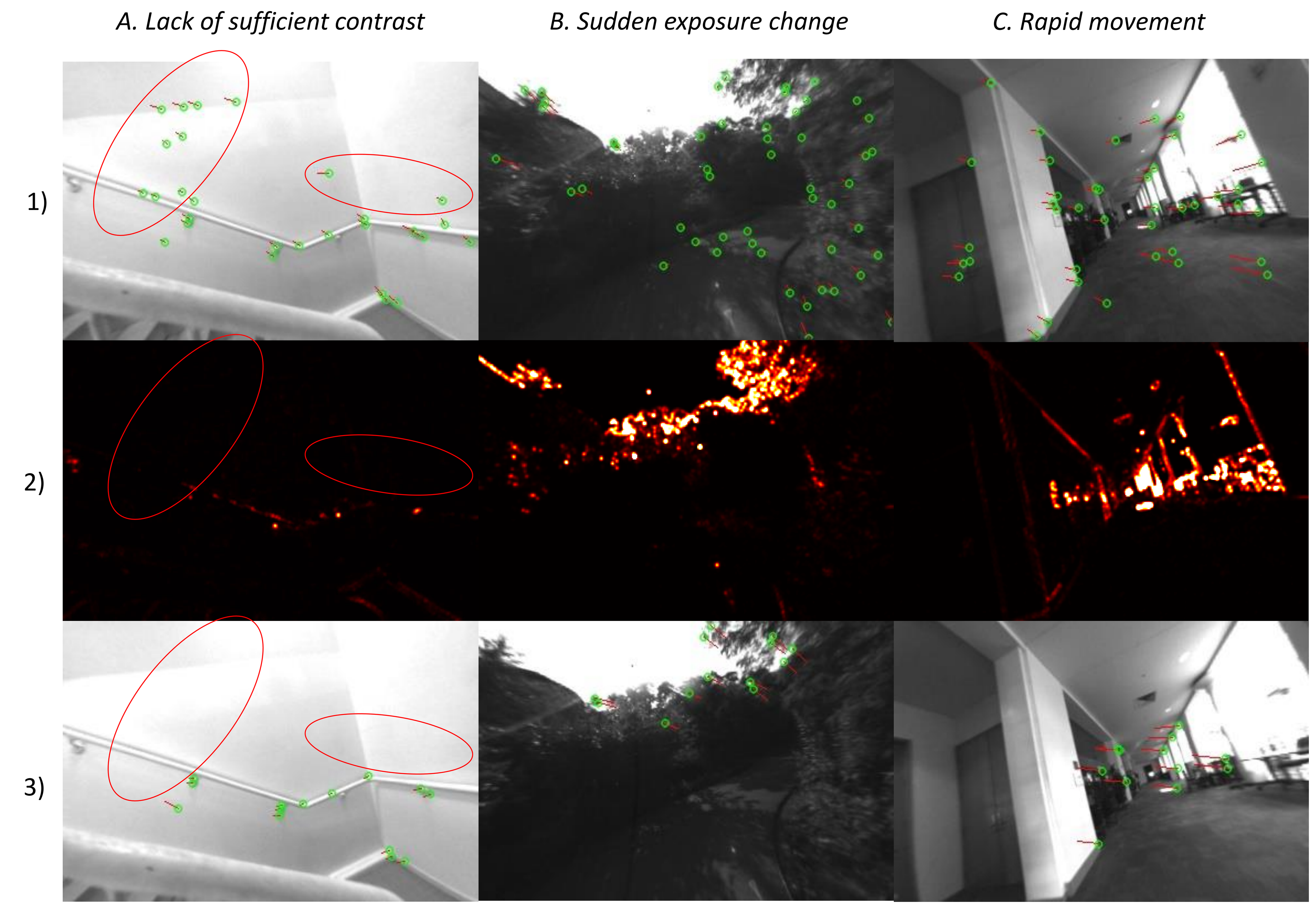
Example of a failure modality: sudden exposure change eliminates all features being tracked

Analysis of Failure Modalities

Saliency maps: Generated for all images using simplified implementation of Itti center-surround saliency algorithm [3]

- Strong correlation exists between saliency map and count + distribution of robust features that survive failure modalities

Common failure modalities



Tracked features in low-saliency regions do not survive failure modalities while tracked features in high-saliency regions do 1) image at time t *(pre-failure modality), 2) saliency map for image at time* t*, 3) image at time* t+1 *(post-failure modality)*

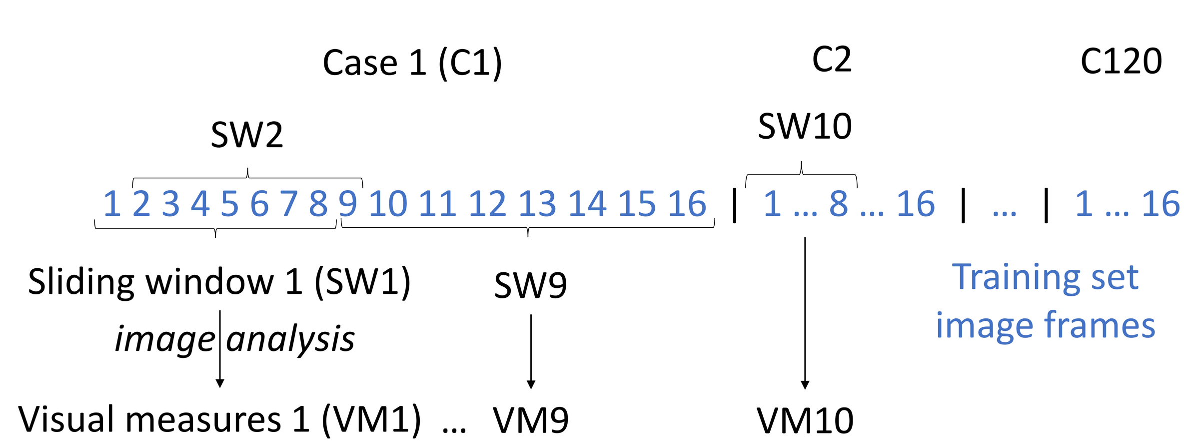
Conclusion: Saliency predicts robustness of tracked features within frame and subsequently predicts feature-based VO failure

Training a Failure Prediction Model

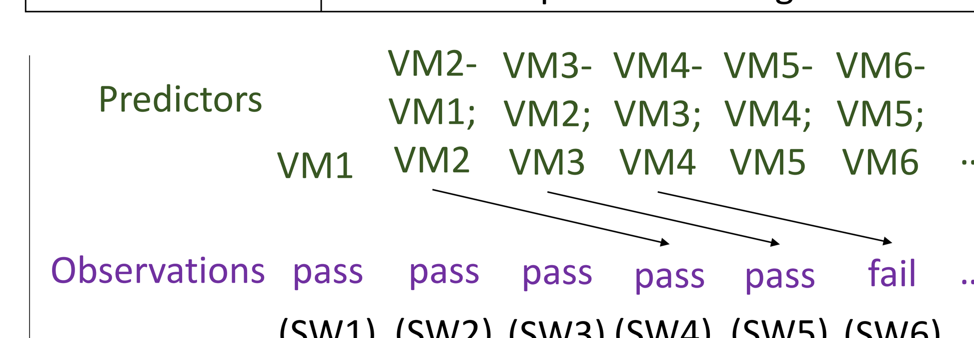
Information extraction:

- Training and testing (cross-validation) sets each consisted of 60 failure and 60 non-failure cases (C) of 16 consecutive frames
- Unit of analysis: sliding window (SW) of 8 most recent images to capture transients (denoted by keyword *change* in table right)
  - Sizes chosen empirically for highest prediction from iteration over a range (C: 3 – 18, SW: 2 – 18)
- Visual measures for sliding windows obtained by accumulating saliency and luminance measures over constituent frames

Relevant Failure Modality	**Visual measures for sliding window**
Lack of sufficient contrast	Total saliency area
	Deviation from average luminance
Exposure change	Change in saliency area
	Change in total intensity
	Histogram of image grids that experienced a significant change in mean intensity
Rapid movement	Distribution of saliency
	Change in center of mass of saliency
Etc.	Average tracked feature count in 3 random frames from previous sliding window



Obtaining visual measures for sliding windows



Training regression to predict VO failure using visual measures

Training a failure prediction model

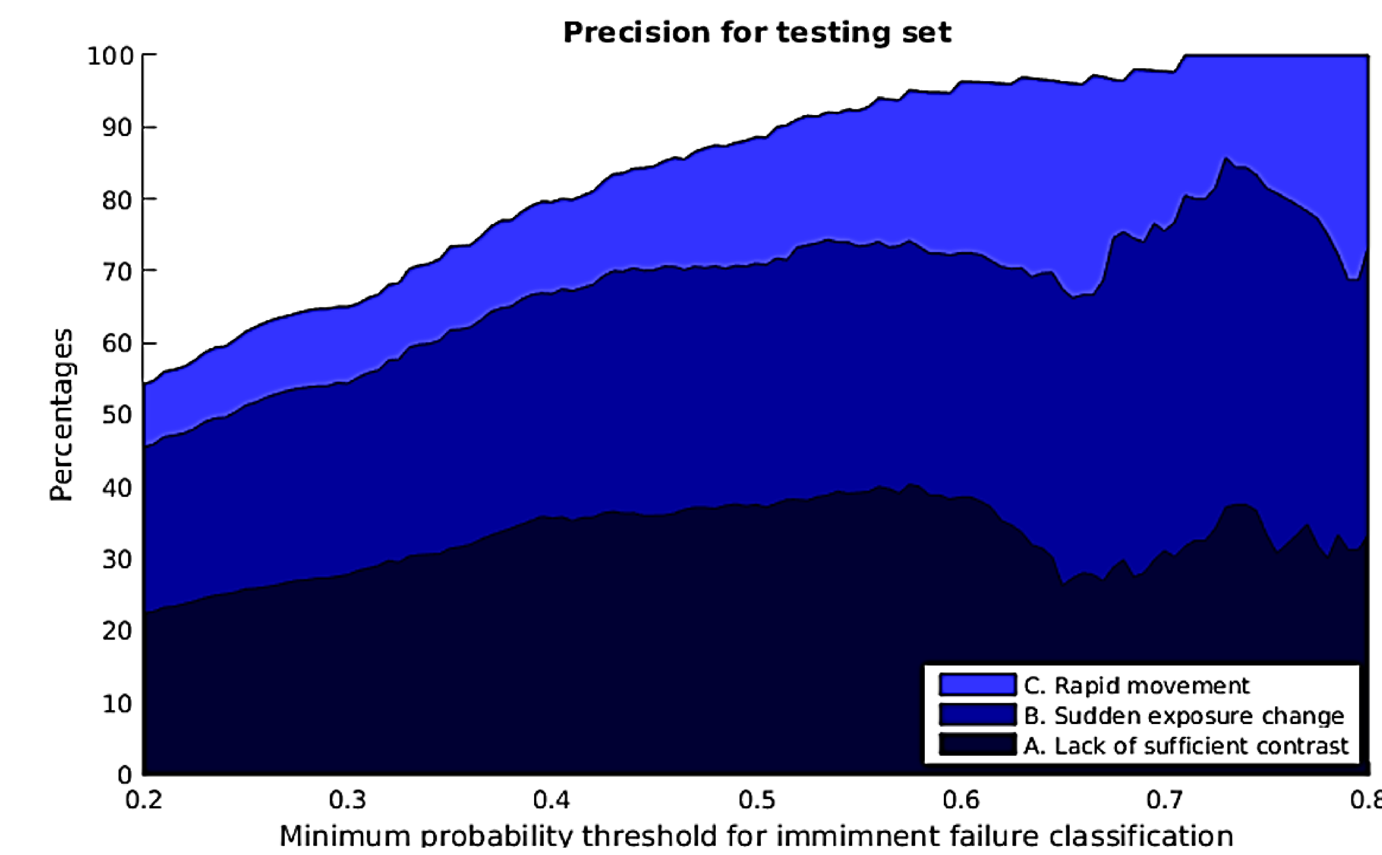
- **Goal**: Predict imminent failure two frames in advance using visual measures at times *t* and *t-1*
- Failure prediction model trained offline with a random forest regression using training dataset [4]
  - Number of decision trees: 15; Minimum number of observations per leaf: 20
  - Predictors: 1) visual measures of current SW, 2) difference between visual measures of current and previous SW
  - Observations: success or failure of SW two images frames away (no features tracked between two consecutive constituent frames)

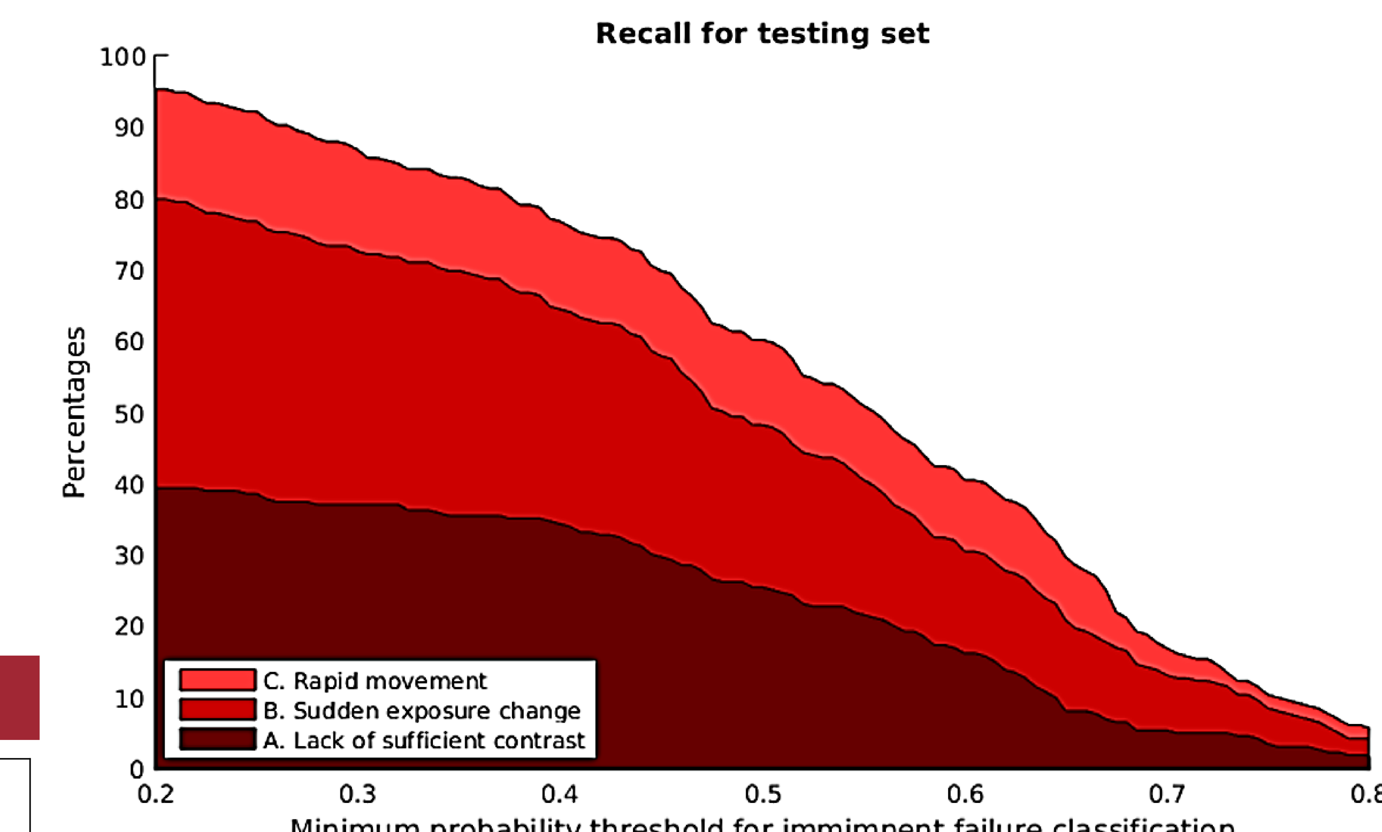
Testing and Results

Testing the failure prediction model

- Predictors computed for training and testing sets, and fed into prediction model
- Probability of failure predicted by model two frames in advance is classified as success or failure with varying threshold

Precision: % of classified failures that were true failures
Recall: % of true failures that were classified as failures





	Training set	Testing set
Max precision	100%	100%
Min precision	54.94%	54.29%
Max recall	99.22%	95.37%
Min recall	13.18%	5.79%

Discussion and limitations

- Comparable results of training and testing sets suggests robustness of failure prediction model in considered environments
- Individual failure modality constituencies (A,B,C) of precision and recall curves roughly match the observation percentages (36%, 47%, 17%), indicating efficacy of visual measure suite at classifying all three considered failure modalities
- High maximum precision and recall at probability thresholds non-inclusive of 0 and 1 suggest that saliency and luminance measures predict imminent failure
- Significant tradeoff between precision and recall indicates current suite is insufficient and additional visual measures are needed, such as interaction and nonlinear combinations of current measures that could capture a possible coupling of failure modes

Future Work

- Explore additional visual measures to increase number of true failure classifications made at higher probability thresholds
- Incorporate failure modality recognition so VO can be augmented appropriately
- Run VO failure prediction and augmentation in real-time, obtain failure prevention results from field tests

References

[1] Itti, I., Koch, C., Niebur, E.:A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254 – 1259 (1998)
[2] Holtz, K., Maturana, D., Scherer, S.,: Learning a Context-Dependent Switching Strategy for Robust Visual Odometry. Submitted to Field and Service Robotics. (2015)
[3] J. Harel, A Saliency Implementation in MATLAB: http://www.klab.caltech.edu/~harel/share/gbvs.php
[4] Breiman, L.: Random Forests. In: Machine Learning 45(1), 5-32 (2001)