Learning to Detect Ground Control Points for Improving the Accuracy of Stereo Matching

Aristotle Spyropoulos[1], Nikos Komodakis[2], Philippos Mordohai[1]
[1] Stevens Institute of Technology, Hoboken, NJ, USA · [2] Ecole Des Ponts, ParisTech, Paris, France

PROBLEM STATEMENT

- Is it possible to predict whether a disparity is right or wrong based on a pixel's features?
- Is it possible to use these predictions to improve the disparity map?

DATASET

The 2003 and 2006 Middlebury College dataset.
27 images in a three-fold cross-validation.

METHOD

- Use features of each image pixel to train a Random Forest classifier to predict whether the assigned disparities are correct.
- Using the same features, test the calculated disparities and generate a prediction of their correctness.
- Use the prediction to select Ground Control Points (GCPs) with very high accuracy and high density.
- Use the GCPs as constraints into an MRF optimizer to improve on the basic disparity maps.

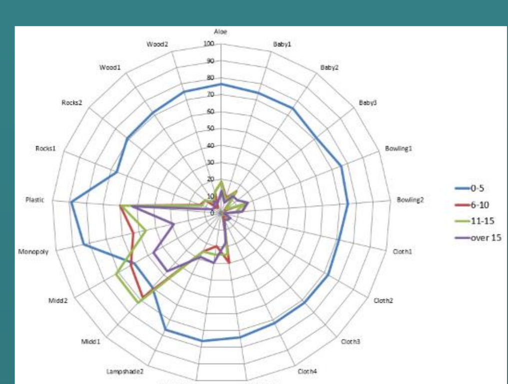
THE FEATURES

1. COST

Negated Normalized Cross Correlation (NCC) in a 5x5 window

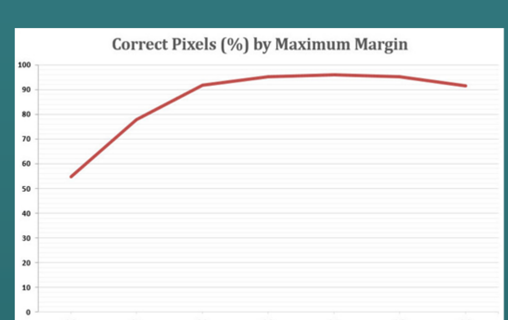
2. DISTANCE FROM IMAGE BORDER

Pixels closer to the image border are more likely to be wrong



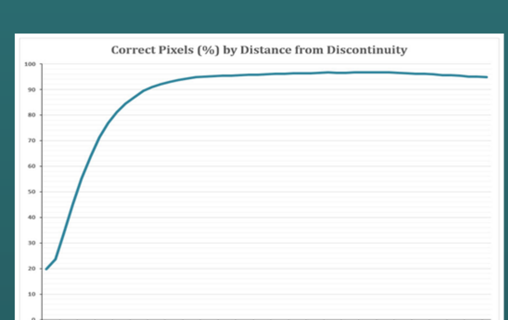
3. MAXIMUM MARGIN

Measures the difference between the two smallest cost values of a pixel



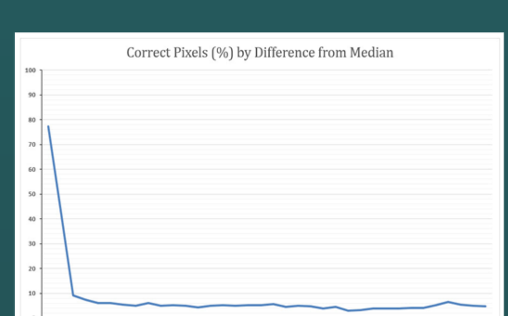
4. DISTANCE FROM DISCONTINUITY

Pixels away from edges of objects are more likely to be correct. We measure DD as the horizontal distance from each pixel to the nearest disparity discontinuity.



5. DIFFERENCE WITH MEDIAN DISPARITY

Difference with median disparity in a 5x5 window



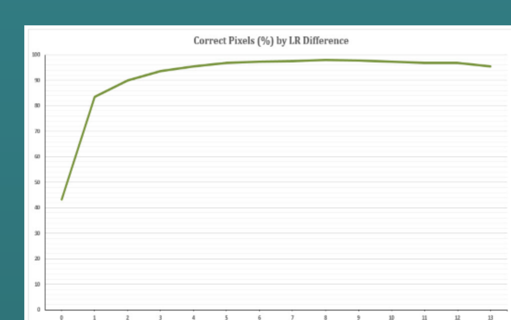
6. LEFT-RIGHT CONSISTENCY

Defined as the absolute difference of the disparities of the left and corresponding right pixel

$$C_{LRC}(x,y) = |d_L(x,y) - d_R(x - d_L, y)|$$

7. LEFT-RIGHT DIFFERENCE

The consistency of the left/right disparities. LRD will be small, if the margin is small, or if the margin is large but the pixel has been mismatched causing a large denominator.

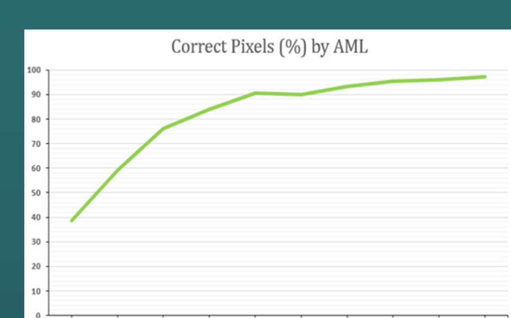


$$C_{LRD}(x,y) = \frac{c_2 - c_1}{|c_1 - min\{c_R(x - d_1, y, d_R)\}|}$$

c_1 = left pixel lowest cost
c_2 = left pixel 2nd lowest cost
c_R = right pixel lowest cost

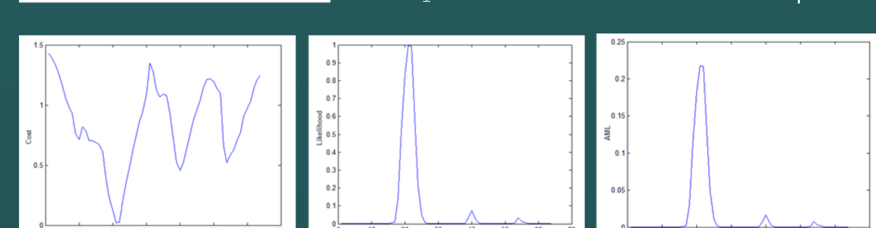
8. ATTAINABLE MAXIMUM LIKELIHOOD

Models the cost of a particular pixel using a Gaussian distribution centered at the minimum cost value for that pixel, c_1



$$C_{\mathrm{AML}} = \frac{e^{-\frac{(c_1 - c_1)^2}{2\sigma_{AML}^2}}}{\sum_d e^{-\frac{(c(d) - c_1)^2}{2\sigma_{AML}^2}}}$$

where:
σ_{AML} = 0.20
$c(d)$ = cost at the calculated disparity
c_1 = smallest cost for that pixel



Cost · Likelihood · AML

RANDOM FOREST

RF was trained using three-fold cross-validation to predict whether an assigned disparity is correct or not

Training Dataset
- Images: 18
- Features: 8
- Pixels: ~2,800,000
- Trees: 50

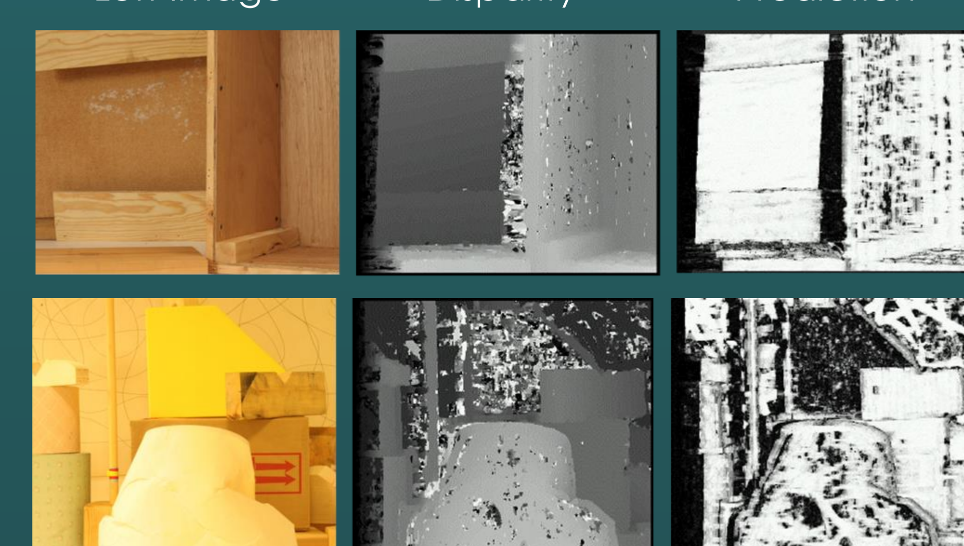
Test Dataset
- Images: 9
- Disparity correctness prediction: 0 = wrong, 1 = correct

The table shows prediction accuracy of non-occluded pixels by using a threshold of 0.5. Columns 2 and 3 correspond to correctly classified pixels; columns 1 and 4 correspond to misclassifications.

	Correct Disparity		Incorrect Disparity	
Image	Y < 0.5	Y ≥ 0.5	Y < 0.5	Y ≥ 0.5
Aloe	4,377	106,143	16,113	5,805
Baby1	1,934	119,735	10,074	3,210
Books	7,612	108,181	21,335	8,824
Cloth1	554	130,283	5,993	174
Lampshade1	9,539	82,016	33,005	8,847
Lampshade2	7,456	84,364	32,910	7,501
Wood1	3,052	125,435	11,711	3,843
...	...	...	...	...
TOTAL	130,142	2,756,764	601,110	177,227
ACCURACY		95.49%	77.23%	

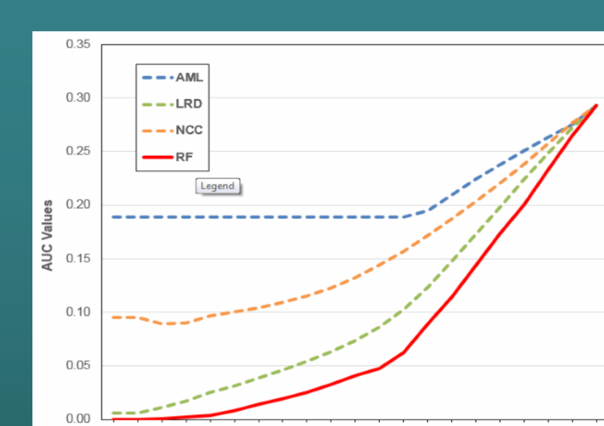
Overall accuracy is 91.6%.

Left Image · Disparity · Prediction

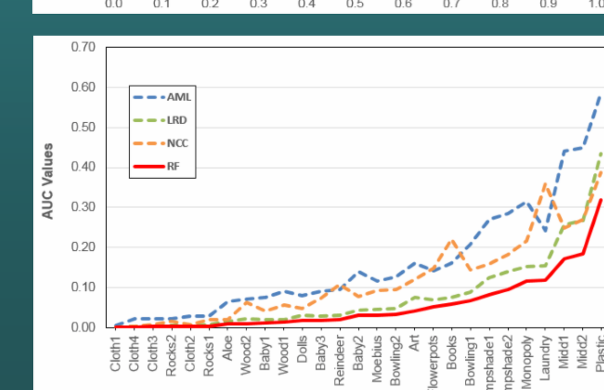


PREDICTION ACCURACY

AUC (Area Under the Curve) values obtained by sorting the disparity assignments according to AML, LRD, NCC and the RF prediction (shown for Bowling1)

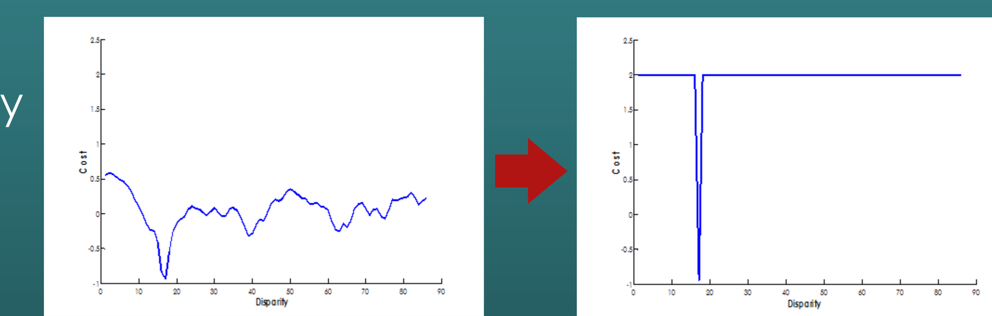


Our method achieves the minimum AUC for every stereo pair



GROUND CONTROL POINTS

- Disparity assignment with higher prediction was selected as a Ground Control Point (GCP). Prediction threshold value was set to 0.7.
- GCPs were not used as hard constraints in MRF. When RF predicted that a given disparity of a pixel was *reliable*, we set the cost of all other disparities to a constant higher value, namely 2, leaving the cost of the selected disparity intact.



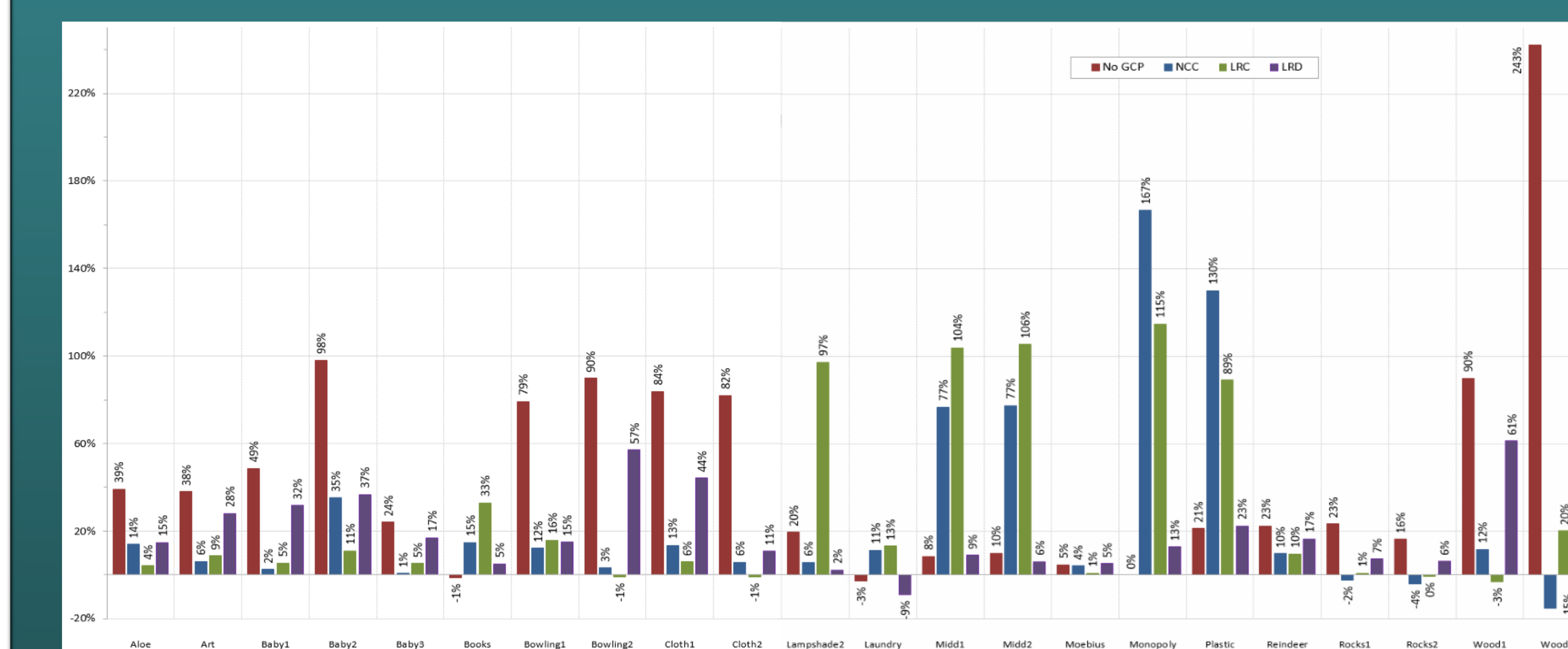
- This allowed MRF to override the GCPs at a higher cost and was proven to be more effective

MRF

$$E = E_{data} + \lambda \cdot E_{smoothness}$$

E_{data} = -NCC modified as above
$E_{smoothness}$ = smoothness term; a Potts model with contrast-weighted edge strength
λ = factor, experimentally set to 2.2

The figure shows error rate differences of RF over other methods and is defined as: $\frac{\varepsilon_{method} - \varepsilon_{RF}}{\varepsilon_{RF}}$

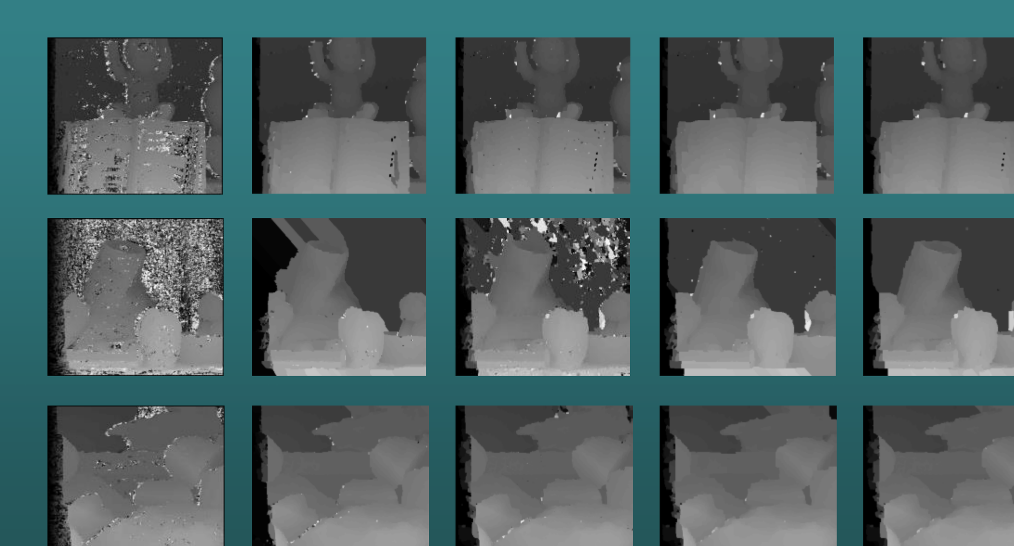


The results by Average Error are:

GCP Type	Average Error
None	9.84%
NCC	9.95%
LRC	10.28%
LRD	8.69%
RF	7.39%

FINAL DISPARITY MAPS

Final disparity maps using an MRF without GCPs (leftmost column) and MRFs with GCPs determined according to NCC, LRC, LRD and the RF predictions (left to right)



Basic MRF · NCC GCPs · LRC GCPs · LRD GCPs · RF GCPs