CBMV: A Coalesced Bidirectional Matching Volume for Disparity Estimation

Konstantinos Batsos, Changjiang Cai, Philippos Mordohai

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Code is available at:
https://github.com/kbatsos/CBMV

Motivation

- Purely data-driven learning-based methods have been successful in stereo matching by replacing hand-crafted rules with data-driven predictions. However these methods tend to overspecialize in the training domain.
- Conventional matching functions enforce desirable invariances which could have been learned from data.
- With CBMV we aim to generate a robust matching volume able to compete with purely data-driven methods, but also able to generalize much better in different domains.
- We integrate conventional matching methods in a supervised learning framework to achieve robustness.
- We show competitive results with the fast MC-CNN architecture and an improved capability to generalize on Middlebury 2014, KITTI 2012 and 2015, and ETH3D benchmarks using the **same model.**
- **Hypothesis:** This is due to the learning method not being directly exposed to image appearance

Method

- Four matchers are coalesced with bidirectional confidence features to create CBMV
- We use a random forest classifier to predict matching likelihood
- The final disparity maps are extracted from the matching volume using the optimization and filtering pipeline of MC-CNN

Features

- The cost computed by each of the four different matchers

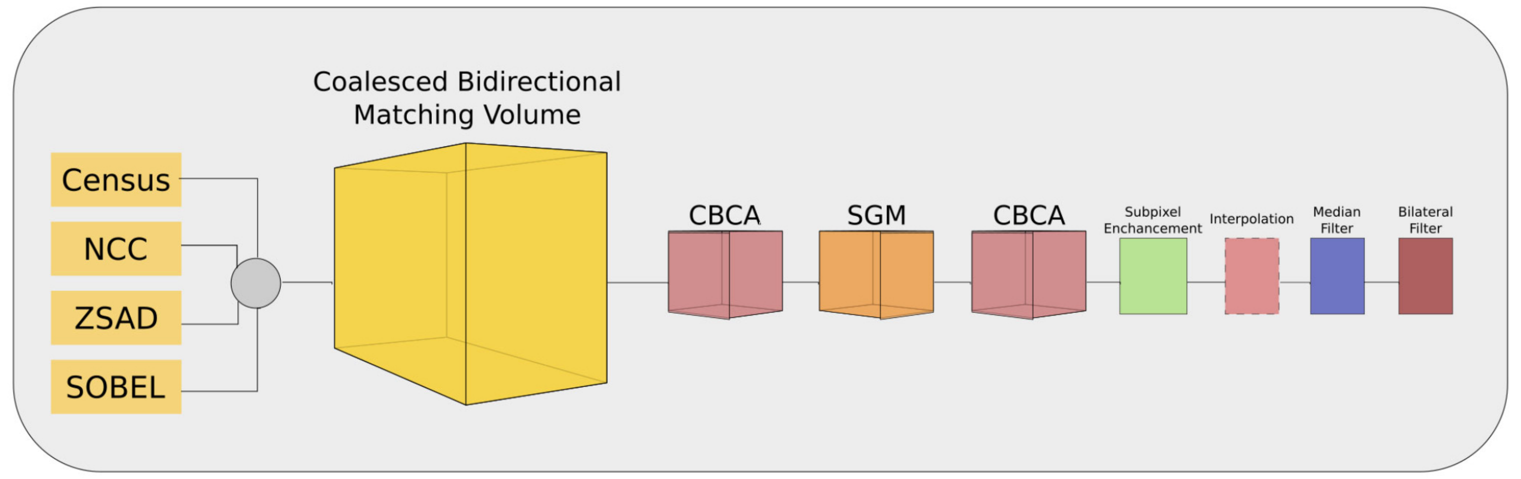
$$C_{cen}(x_L, d) = CENSUS(x_L, d)$$

- For each matching function we compute two features, Ratio and Likelihood in a bidirectional manner

Ratio
$$R^L_{cen}(x_L, d) = \frac{c^L_{cen,min}}{C_{cen}(x_L, d)}$$

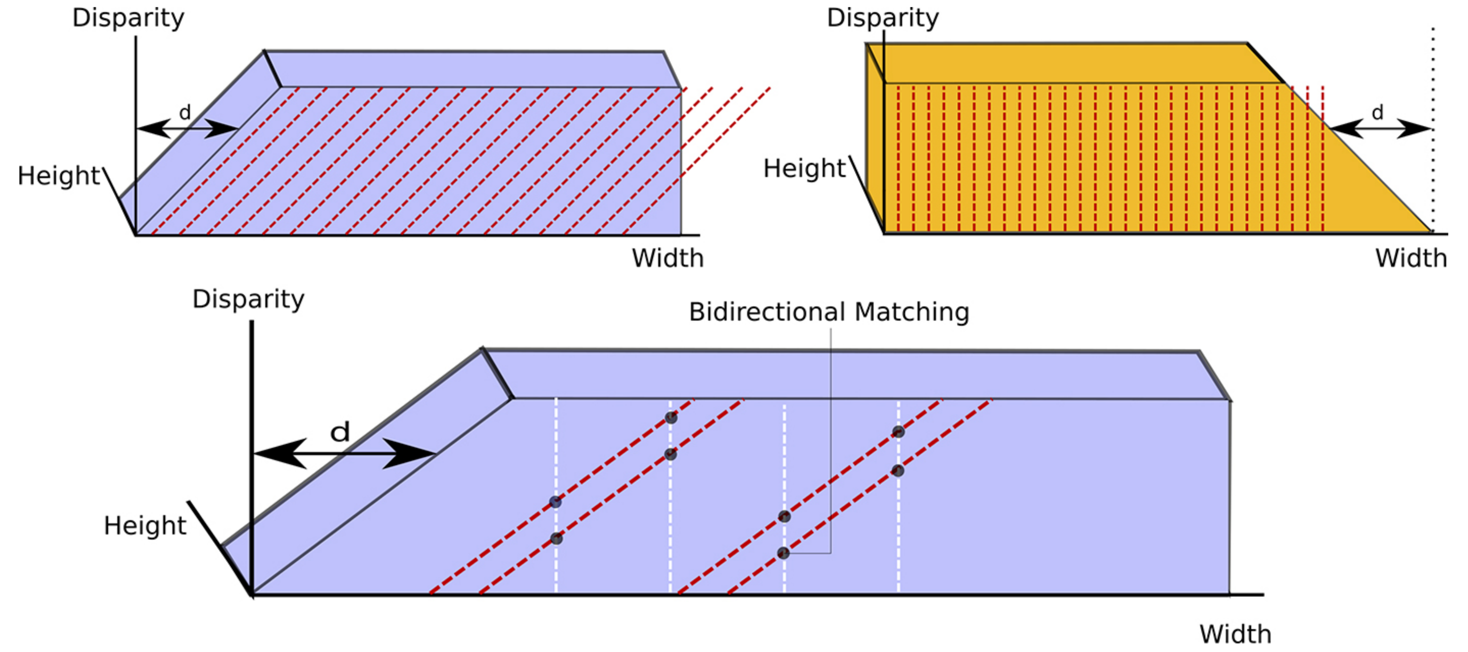
Likelihood
$$L^L_{cen}(x_L, d) = \frac{e^{-\frac{(C_{cen}(x_L,d) - c^L_{cen,min})^2}{2\sigma^2}}}{\sum_i e^{-\frac{(C_{cen}(x_L,d_i) - c^L_{cen,min})^2}{2\sigma^2}}}$$

Stereo Pipeline



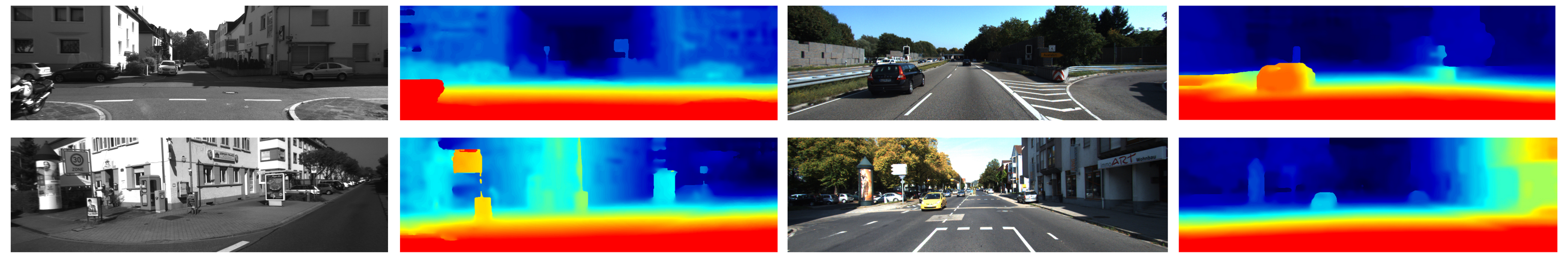
Bidirectional Matching

- For a given element of the matching volume, the ratio and likelihood features are computed along the yellow and red lines corresponding to the right and left epipolar lines respectively.
- Black dots denote a few intersections of left and right epipolar lines on the matching volume. Each intersection is a matching hypothesis linking a pixel in the left image with a pixel in the right image.



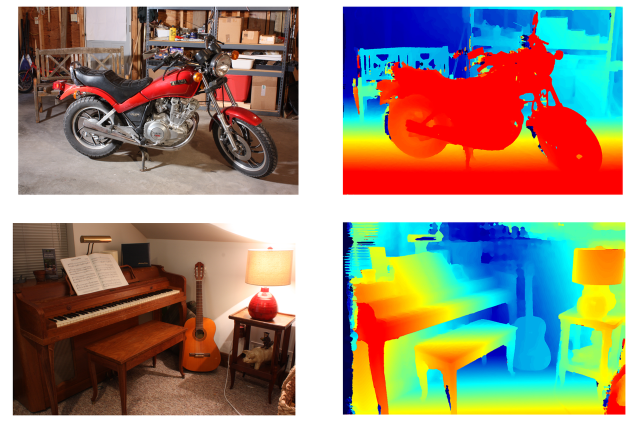
Training

CBMV is trained on the training and additional sets of the Middlebury 2014 dataset for **ALL** benchmark submissions



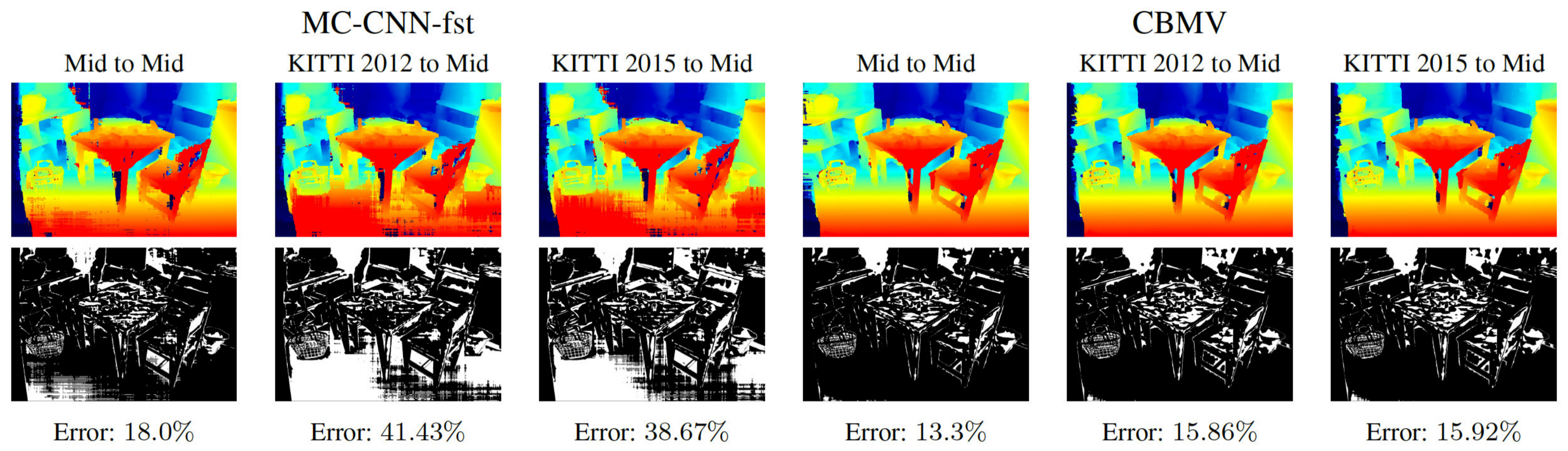
Results

	bad-2.0 nonocc	bad-2.0 all	avgerror all	rms-error all
Middlebury 2014 test set				
MC-CNN-acrt	**8.08%**	**19.1%**	17.9	55.0
CBMV(ours)	11.1%	20.5%	**14.4**	**46.9**
MC-CNN-fst	9.47%	20.6%	19.3	55.7



Generalization

		Test set KITTI 2012 (Out-Noc)			KITTI 2015 (Out-All)			Middlebury (bad 2.0)		
		MC-ac	MC-fst	CBMV	MC-ac	MC-fst	CBMV	MC-ac	MC-fst	CBMV
Training set	KITTI 2012	0%	0%	0%	23.07%	13.28%	-0.41%	40.20%	33.58%	7.00%
	KITTI 2015	63.98%	17.54%	3.02%	0%	0%	0%	79.39%	41.62%	7.69%
	Middlebury	17.62%	10.51%	-4.62%	38.15%	18.79%	-2.09%	0%	0%	0%



Robust Vision Challenge

For our robust vision challenge submission, we replaced the post processing and optimization pipeline with the local expansion algorithm [*T. Taniai et. al. Continuous 3D label stereo matching using local expansion moves*]. Our results show that CBMV **does not underfit or overfit** any of the benchmarks.

dataset	Middlebury		KITTI 2015		ETH3D	
metric	bad 2.0 noc	bad 2.0 all	bad 3.0 noc	bad 3.0 all	bad 1.0 noc	bad 1.0 all
CBMV_ROB	7.65%	13.3%	4.52%	4.97%	4.14%	4.66%
CBMV	11.1%	20.5%	4.58%	5.06%	5.35%	5.97%

Acknowledgements

This research has been supported by the National Science Foundation under awards IIS 1527294 and 1637761. We are grateful to Enrique Dunn for constructive discussions.