Supplementary material for paper: Learning to predict stereo reliability enforcing local consistency of confidence maps

Matteo Poggi, Stefano Mattoccia

University of Bologna
Department of Computer Science and Engineering (DISI)
Viale del Risorgimento 2, Bologna, Italy
matteo.poggi8@unibo.it,stefano.mattoccia@unibo.it

This document provides detailed formulation of the confidence measures and additional experimental results concerned with paper "Learning to predict stereo reliability enforcing local consistency of confidence maps", CVPR 2017. Single column format is adopted to improve readability. This document is organized into two main sections. Section 1 lists the confidence measures analyzed in our paper, grouping them into stand-alone and machine-learning based, Section 2 shows for two images additional examples of confidence maps obtained by each original confidence measure and by its *plus* counterpart, referred to as +, proposed in the paper.

1 Confidence measures

We report the exhaustive list of state-of-the-art confidence measures, grouped into *stand-alone* and *machine-learning* based measures, considered in our paper.

1.1 Stand-alone measures

We recall the definition of each stand-alone confidence measure according to literature. As shown in Figure 1, given a pixel $\mathbf{p} = (p_x, p_y)$, we will refer to its minimum cost as $c_1(\mathbf{p})$, the second minimum as $c_2(\mathbf{p})$ and the second local minimum as $c_{2m}(\mathbf{p})$. We refer to a matching cost for any disparity hypothesis d as $c_d(\mathbf{p})$ The disparity hypothesis corresponding to $c_1(\mathbf{p})$ will be referred to as $d_1(\mathbf{p})$, the one to $c_2(\mathbf{p})$ as $d_2(\mathbf{p})$ and so on. If not specified, costs and disparities refers to left image pixels (L). When talking about right image (R), we introduce the ^{R} notations on both costs (e.g., $c_1^R(\mathbf{p})$) and disparities. We denote as $\mathbf{p'} = (p'_x, p'_y)$ the matching pixel for $\mathbf{p}$ according to d_1 (i.e., $p'_x = p_x + d_1(\mathbf{p})$, $p'_y = p_y$). Finally, we denote with ^{LL} matching costs and disparities related to self-matching stereo on the left image (i.e., stereo algorithm processing the left image as both reference and target).

- **PKRN** (Peak Ratio Naive), reviewed in [1]

$$C_{PKRN}(\mathbf{p}) = \frac{c_2(\mathbf{p})}{c_1(\mathbf{p})} \tag{1}$$

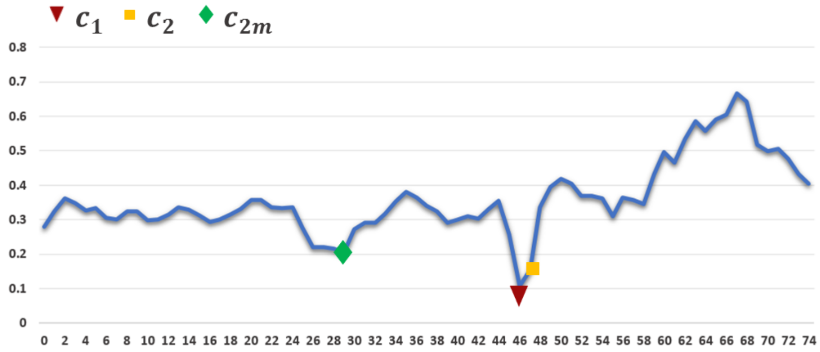


Fig. 1. Example of cost curve, showing the matching cost c_1, the second minimum c_2 and the second local minimum c_{2m}. In this example, the disparity d_1 is equal to 46.

– **PKR** (Peak Ratio), reviewed in [1]

$$C_{PKR}(\mathbf{p}) = \frac{c_{2m}(\mathbf{p})}{c_1(\mathbf{p})} \tag{2}$$

– **MSM** (Matching Score Measure), reviewed in [1]

$$C_{MSM}(\mathbf{p}) = -c_1(\mathbf{p}) \tag{3}$$

– **MMN** (Maximum Margin), reviewed in [1]

$$C_{MMN}(\mathbf{p}) = c_2(\mathbf{p}) - c_1(\mathbf{p}) \tag{4}$$

– **WMN** (Winner Margin), reviewed in [1]

$$C_{WMN}(\mathbf{p}) = \frac{c_{2m}(\mathbf{p}) - c_1(\mathbf{p})}{\sum_d c_d(\mathbf{p})} \tag{5}$$

– **MLM** (Maximum Likelihood Measure), reviewed in [1]

$$C_{MLM}(\mathbf{p}) = \frac{e^{-\frac{c_1(\mathbf{p})}{2\sigma^2_{MLM}}}}{e^{-\frac{c_d(\mathbf{p})}{2\sigma^2_{MLM}}}} \tag{6}$$

– **PER** (Perturbation), proposed in [2] and adopted in [3]

$$C_{PER}(\mathbf{p}) = \sum_{d \neq d_1} e^{-\frac{(c_1(\mathbf{p}) - c_d(\mathbf{p}))^2}{s^2}} \tag{7}$$

– **NEM** (Negative Entropy Measure), reviewed in [1]

$$C_{NEM} = -\sum_d p(d) \log p(d)$$
$$p(d) = \frac{e^{-c_1}}{\sum_d e^{-c_d(\mathbf{p})}} \tag{8}$$

- **LRD** (Left-Right Difference), reviewed in [1]

$$C_{LRD}(\mathbf{p}) = \frac{c_2(\mathbf{p}) - c_1(\mathbf{p})}{|c_1(\mathbf{p}) - \min_d c^R_d(\mathbf{p}'))|} \tag{9}$$

- **CUR** (Curvature), reviewed in [1]

$$C_{CUR} = -2c_1(\mathbf{p}) + c_{d_1-1}(\mathbf{p}) + c_{d_1+1}(\mathbf{p}) \tag{10}$$

- **DSM** (Distinctiveness Similarity Measure), reviewed in [1]

$$C_{DSM} = \frac{C_{L\_DTS}(\mathbf{p}) \times C_{R\_DTS}(\mathbf{p}')}{c_1^2(\mathbf{p})}$$
$$C_{L\_DTS} = \min_{d \in d_s, d \neq 0} c^{LL}_d(\mathbf{p})$$
$$C_{R\_DTS} = \min_{d \in d_s, d \neq 0} c^{RR}_d(\mathbf{p}) \tag{11}$$

- **AML** (Attainable Maximum Likelihood), reviewed in [1]

$$C_{AML} = \frac{1}{e^{-\frac{c_d(\mathbf{p})}{2\sigma^2_{AML}}}} \tag{12}$$

- **NOI** (Number Of Inflections), reviewed in [1]

$$C_{NOI} = |M|$$
$$M = \{d_i : c_{d_i-1}(\mathbf{p}) > c_{d_i}(\mathbf{p}) \wedge c_{d_i}(\mathbf{p}) < c_{d_i+1}(\mathbf{p})\} \tag{13}$$

- **SAMM** (Self-Aware Matching Measure), proposed in [4] and reviewed in [1]

$$C_{SAMM} = \frac{\sum_d (c_{d-d_1}(\mathbf{p}) - \mu) - (c^{LL}_{d-d_1}(\mathbf{p}) - \mu_{LL})}{\sigma \sigma_{LL}} \tag{14}$$

- **WMNN** (Winner Margin Naive), reviewed in [1]

$$C_{WMNN} = \frac{c_2(\mathbf{p}) - c_1(\mathbf{p})}{\sum_d c_d(\mathbf{p})} \tag{15}$$

- **LRC** (Left-Right Consistency), reviewed in [1]

$$C_{LRC}(x, y) = -|d_1 - d_R(\mathbf{p}')| \tag{16}$$

- **LC** (Local Curve), proposed in [5]

$$C_{LC} = \frac{\max\left(c_{d_1-1}(\mathbf{p}), c_{d_1+1}(\mathbf{p})\right) - c_1(\mathbf{p})}{\gamma} \tag{17}$$

- **UC** (Uniqueness Constraint), proposed in [6]

$$UC(\mathbf{p}) = \begin{cases} 0, \text{ if } d_1(\mathbf{p}) \neq d_1^R(p - d_1(\mathbf{p})) \text{ and } c_1 \neq \min_{q \in Q} c_1(q) \\ 1, \text{ otherwise} \end{cases} \tag{18}$$

  being Q the set of pixels matching the same pixel on the right image

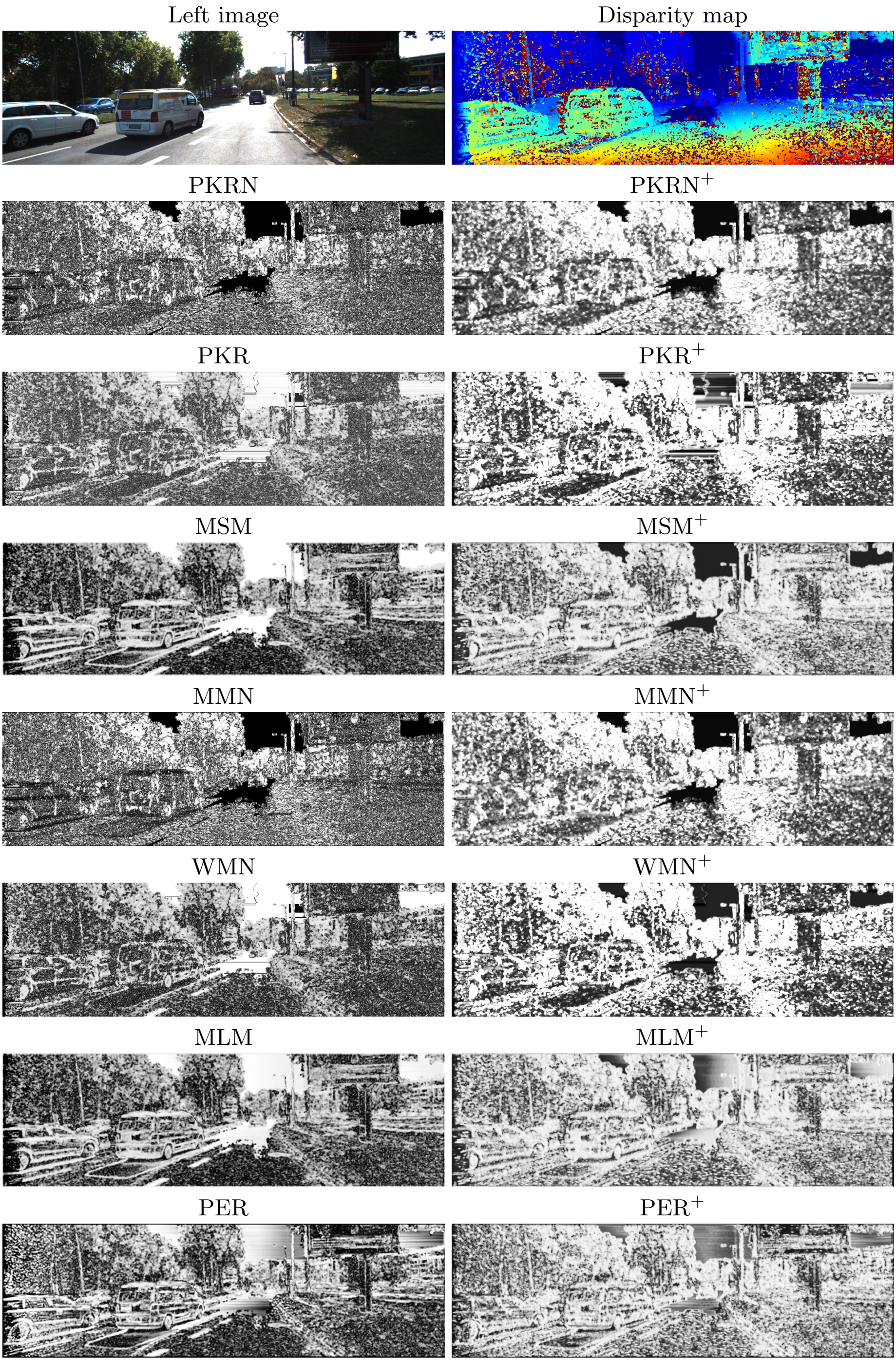
1.2 Machine-learning based measures

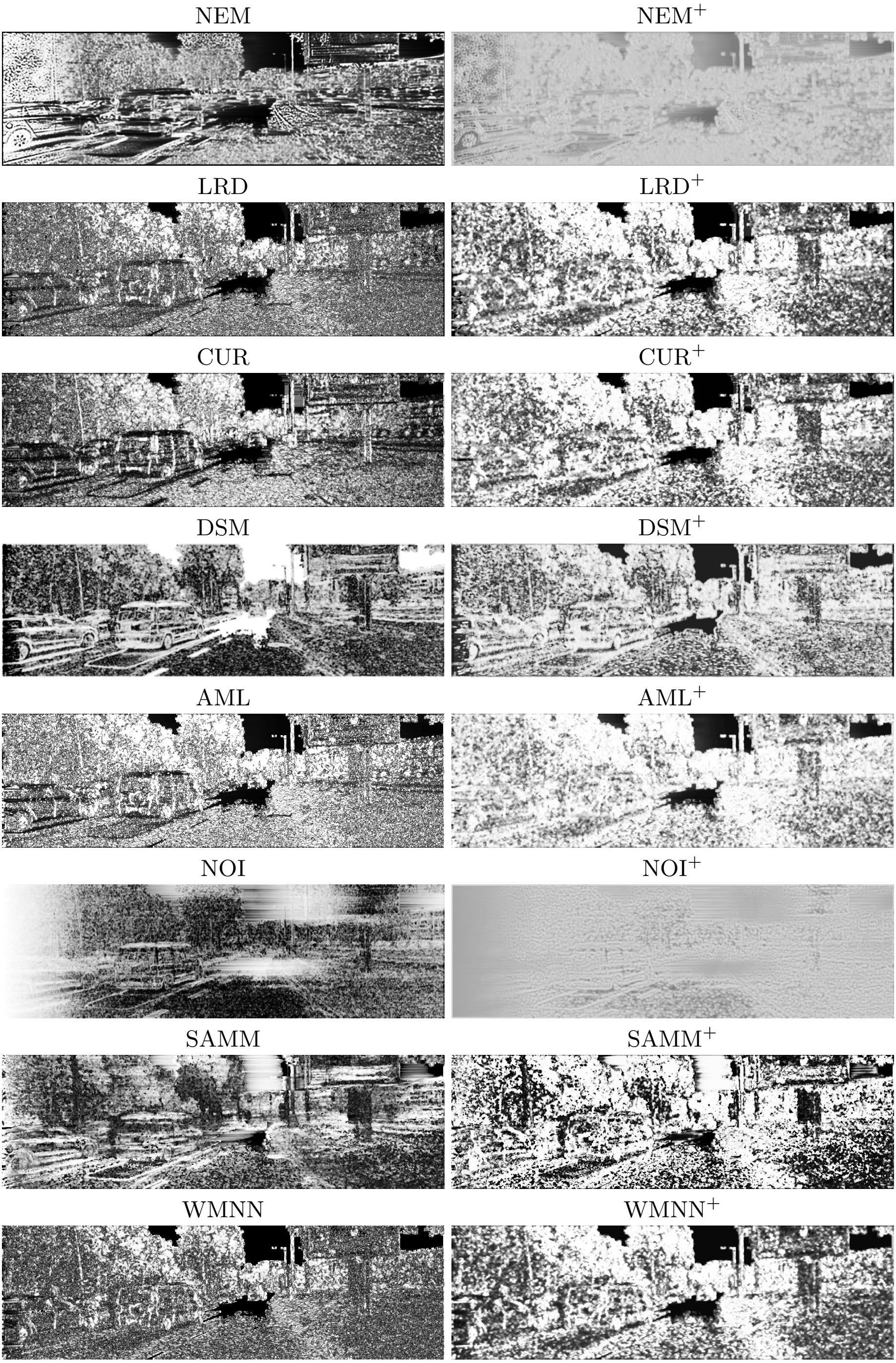
We briefly review the five machine-learning measures considered in the paper.

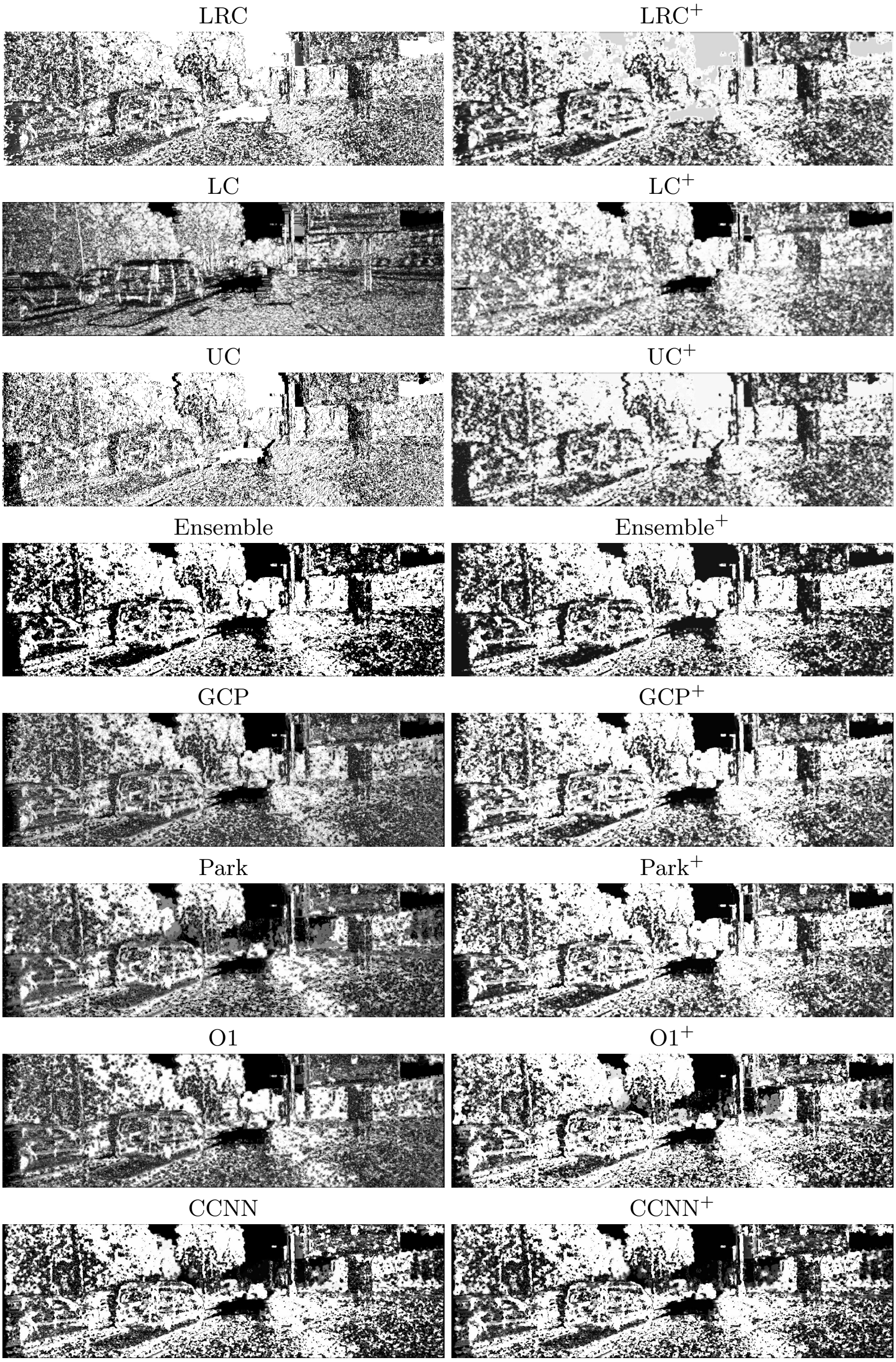
- **Ensemble**, proposed in [3]. It combines 23 confidence measures (some computed at multiple resolutions) fed to a random forest trained in classification mode.
- **GCP**, proposed in [7]. It combines 5 confidence measures (MSM, MMN, AML, LRC, LRD) and three additional features (i.e., *distance to border*, *distance to discontinuities* and *median deviation of disparity*), fed to a random forest trained in regression mode.
- **Park**, proposed in [8]. It combines PKR, PKRN, MSM, MMN, WMN, MLM, NEM, LRD, CUR, LRC and PER confidence measures plus *distance to border*, *distance to edges*, *horizontal gradient magnitude*, *median deviation* and *variance of disparity* (the latter two computed on 5×5, 7×7, 9×9 and 11×11 patches) in a vector of 22 features, fed to a random forest trained in regression mode.
- **O1**, proposed in [9]. It relies on a vector of 20 features extracted from the disparity map fed to a random forest trained in regression mode. The features are: *disparity agreement*, *disparity scattering*, *median*, *variance* and *median deviation* of disparity (the latter three computed on 5×5, 7×7, 9×9 and 11×11 patches).
- **CCNN**, proposed in [10]. The confidence map is inferred from scratch processing the raw disparity map by means of a convolutional neural network with perceptive field of size 9×9 .

2 Confidence maps

In this section, we report a qualitative comparison between confidence maps computed by stand-alone or machine-learning based confidence measures and their proposed *plus* counterpart training our networks, as for any experimental results reported in this document and in the submitted paper, on 25 images of the KITTI 2012 datasets. In most cases, the benefit yielded by our framework can be clearly perceived comparing the original confidence map and its *plus* version. We report detailed experimental results for image #93 from KITTI 2015 dataset and for *Motorcycle* image from Middlebury 2014 dataset.

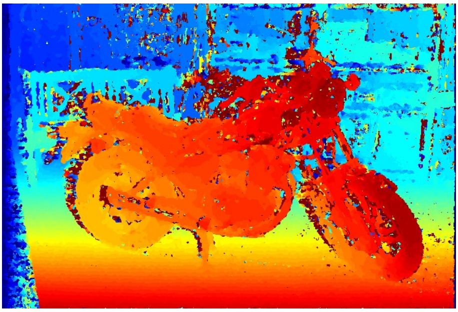

Left image
Disparity map
PKRN
PKRN+
PKR
PKR+
MSM
MSM+
MMN
MMN+
WMN
WMN+
MLM
MLM+
PER
PER+


NEM
NEM$^+$
LRD
LRD$^+$
CUR
CUR$^+$
DSM
DSM$^+$
AML
AML$^+$
NOI
NOI$^+$
SAMM
SAMM$^+$
WMNN
WMNN$^+$

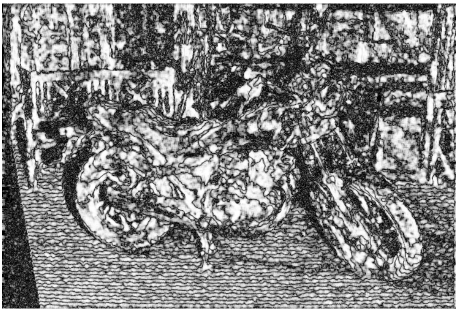

LRC
LRC$^+$
LC
LC$^+$
UC
UC$^+$
Ensemble
Ensemble$^+$
GCP
GCP$^+$
Park
Park$^+$
O1
O1$^+$
CCNN
CCNN$^+$

Left image

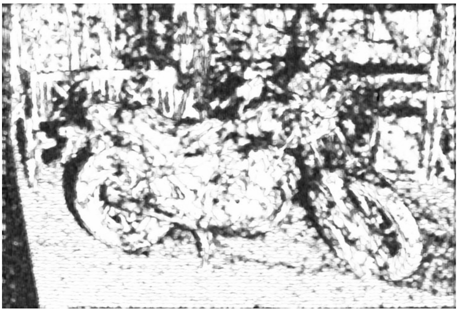
Disparity map



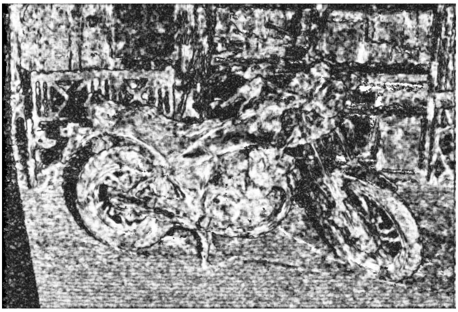
PKRN



PKRN$^+$

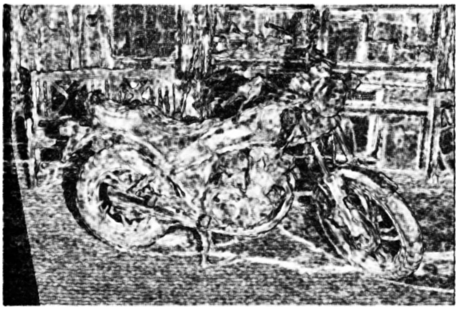


PKR

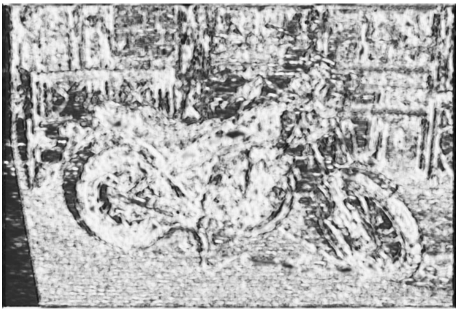


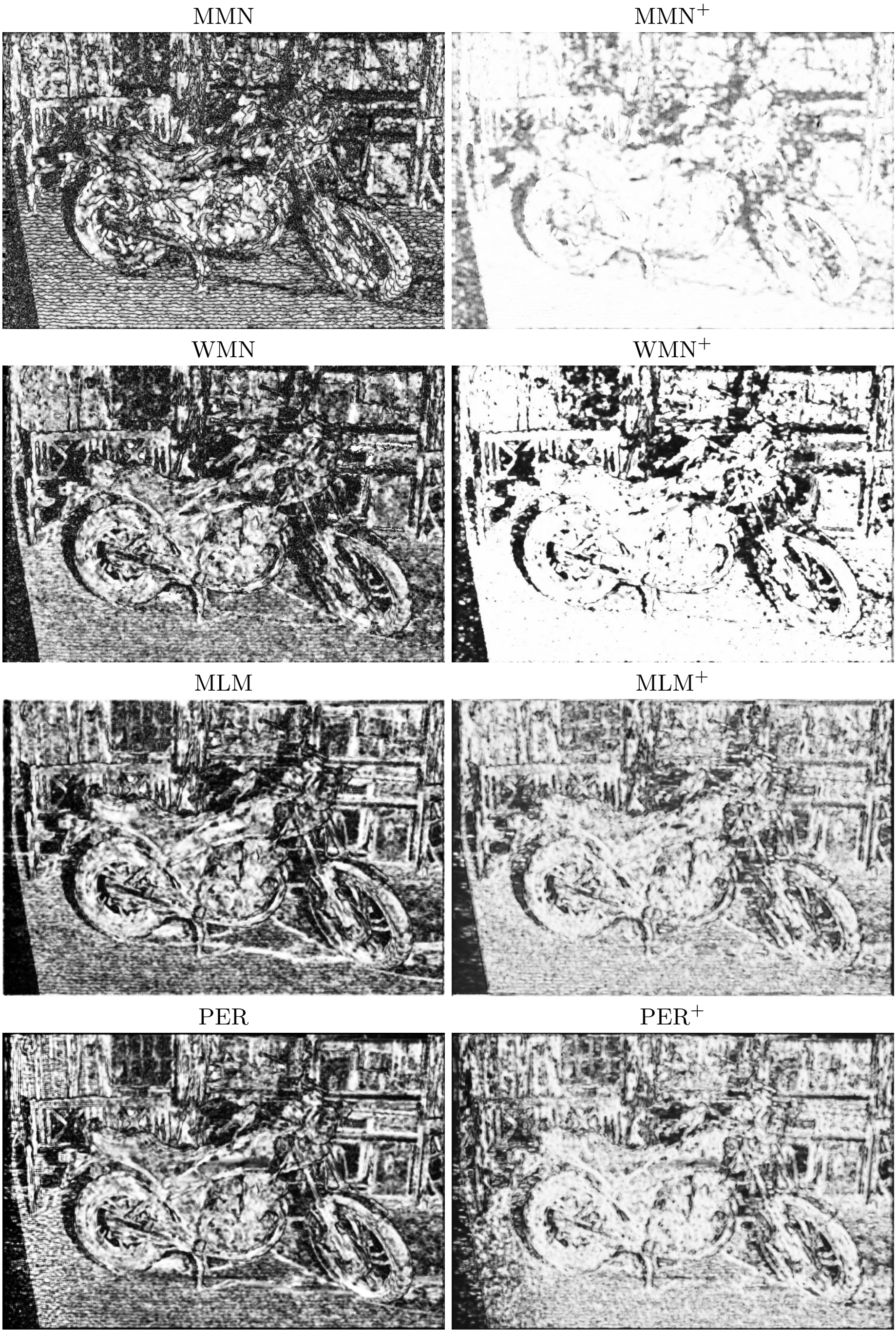
PKR$^+$

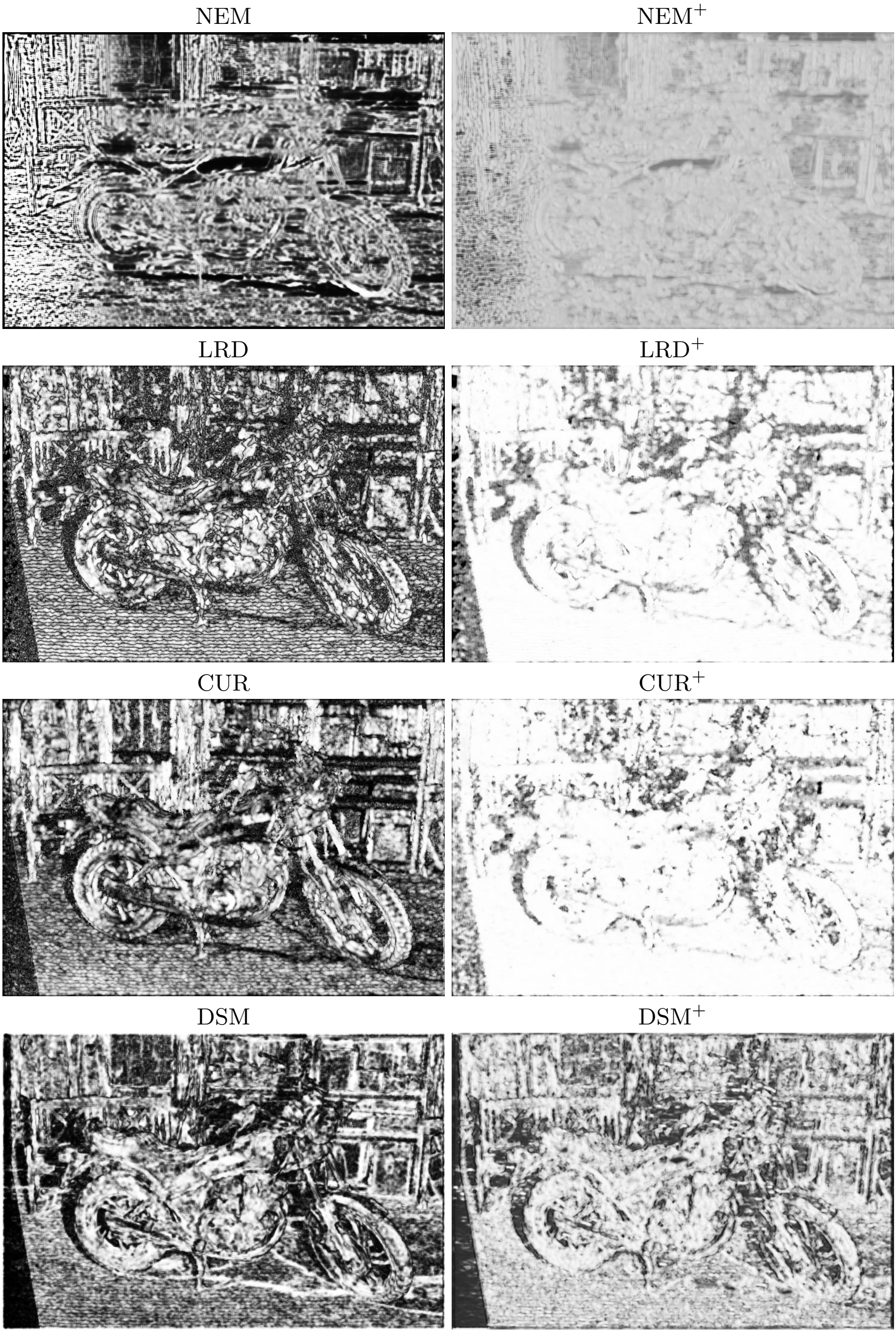
MSM

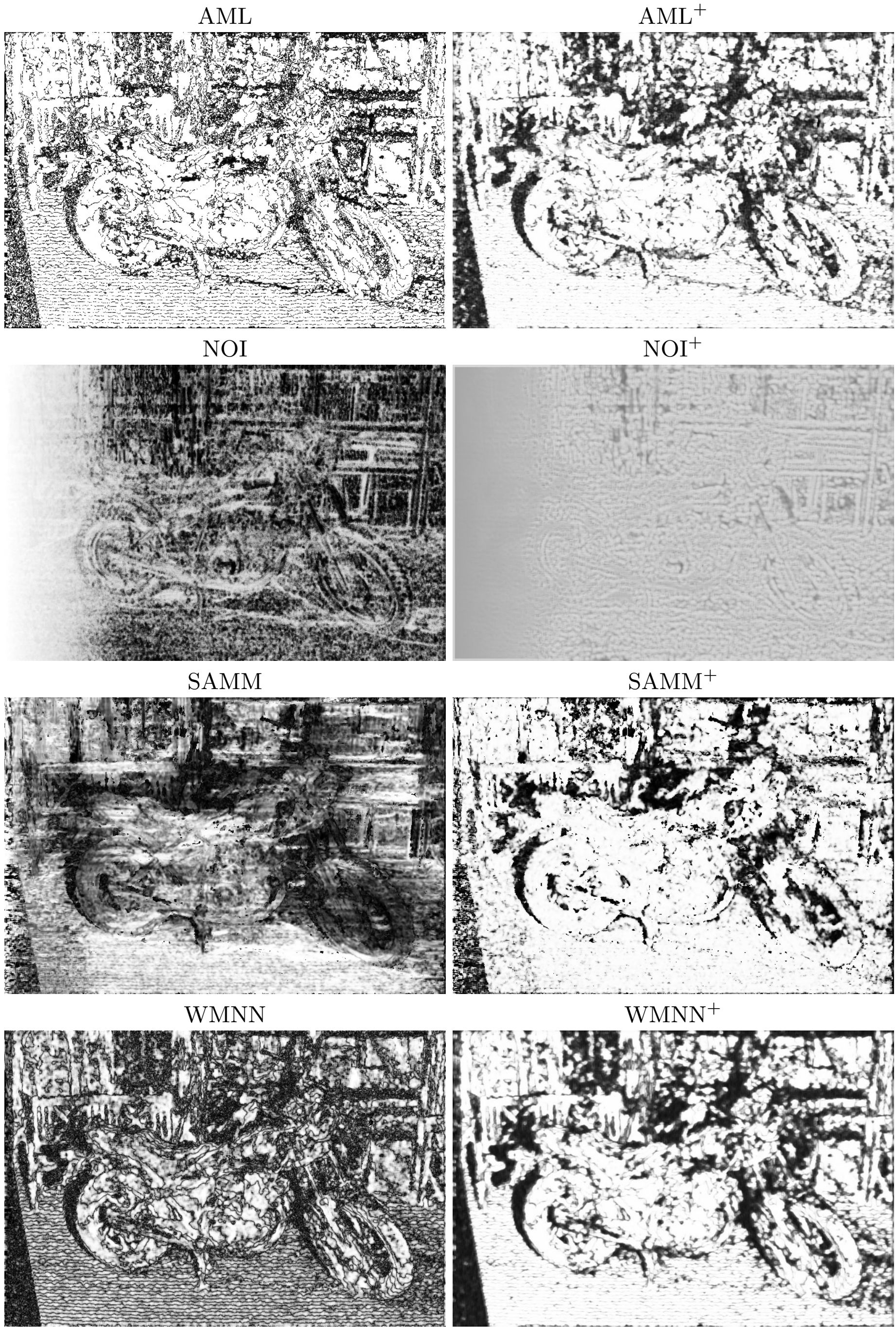


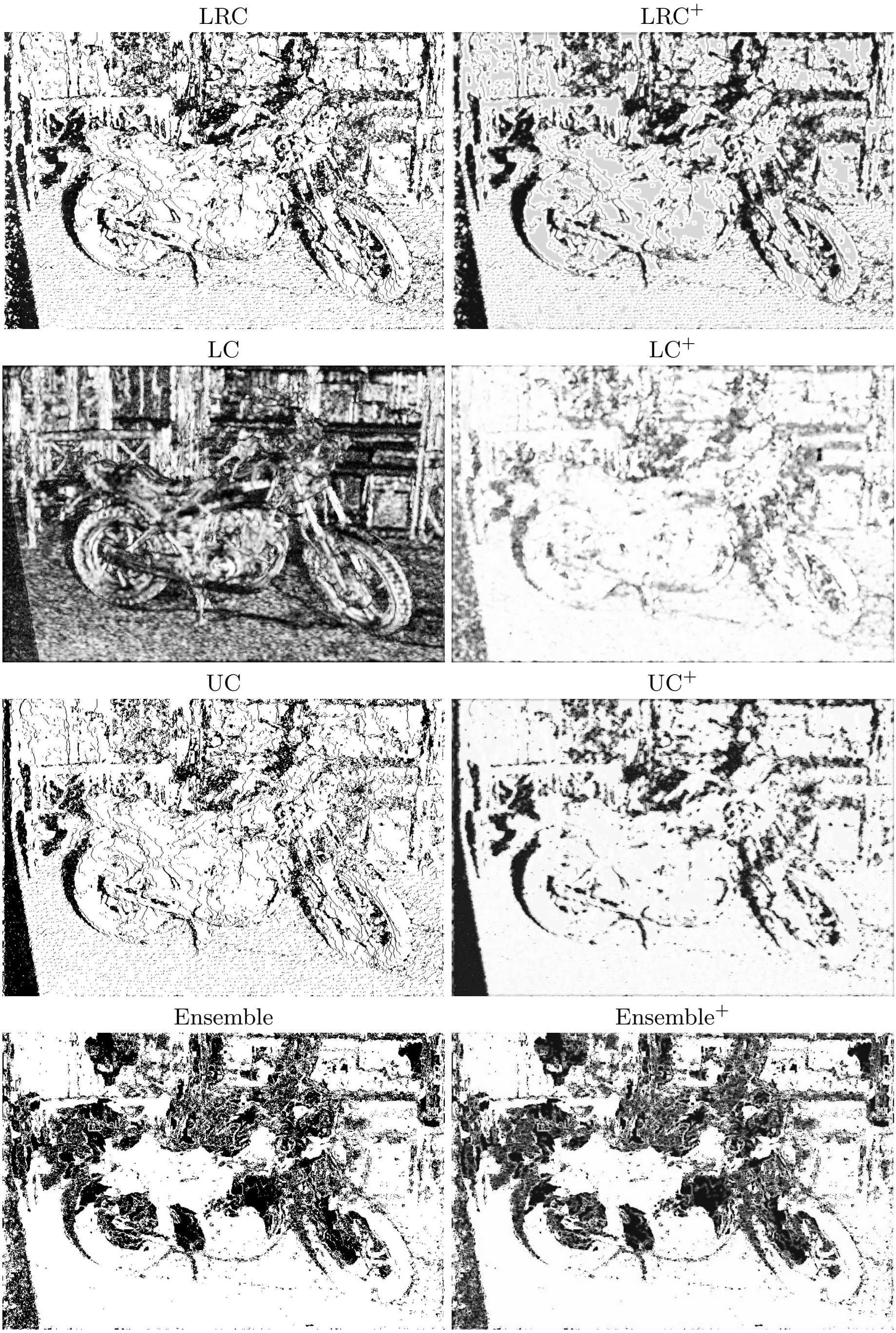
MSM$^+$

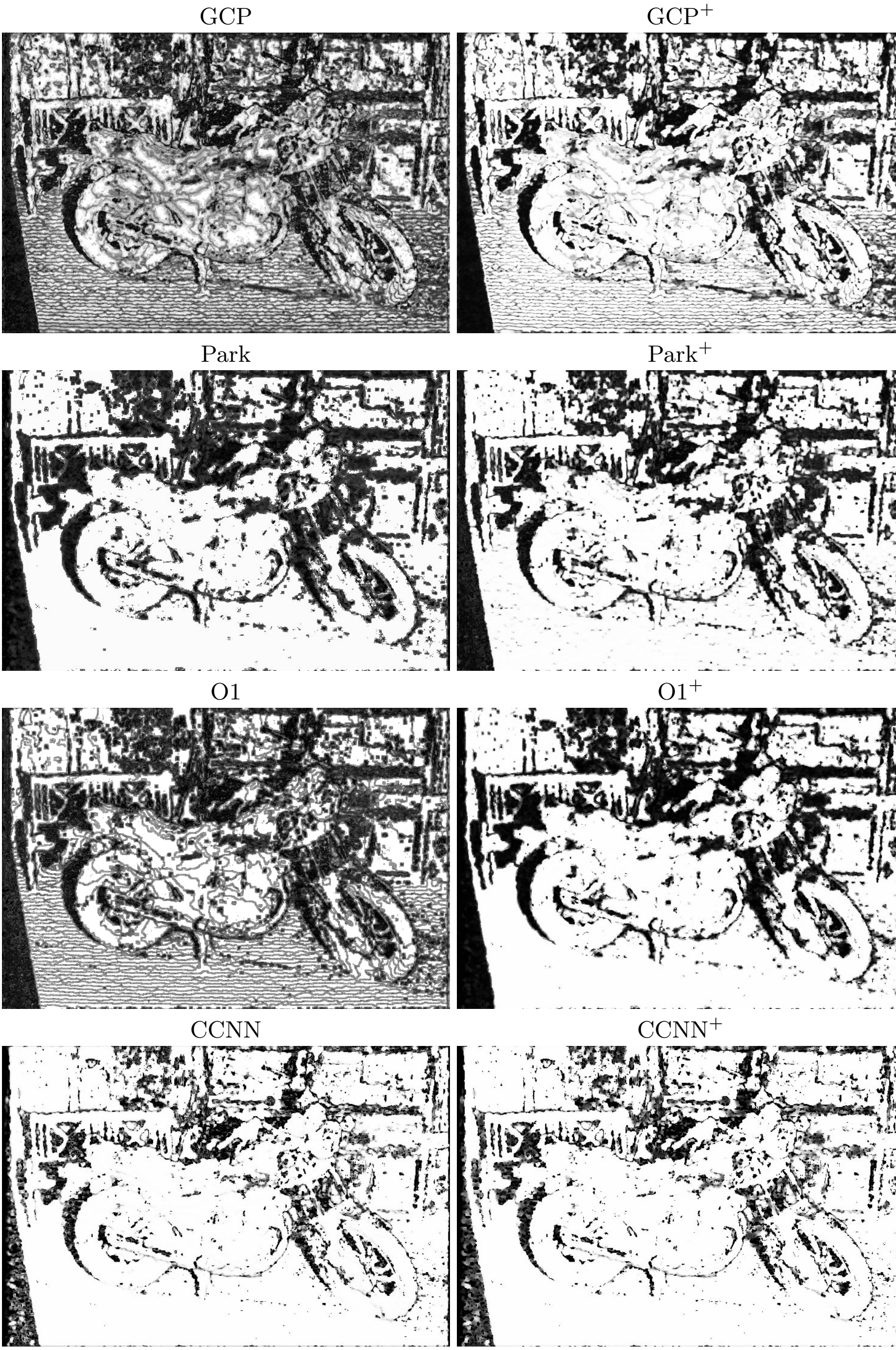



MMN
MMN$^+$
WMN
WMN$^+$
MLM
MLM$^+$
PER
PER$^+$


NEM
NEM$^+$
LRD
LRD$^+$
CUR
CUR$^+$
DSM
DSM$^+$


AML
AML$^+$
NOI
NOI$^+$
SAMM
SAMM$^+$
WMNN
WMNN$^+$


LRC
LRC$^+$
LC
LC$^+$
UC
UC$^+$
Ensemble
Ensemble$^+$


GCP
GCP$^+$
Park
Park$^+$
O1
O1$^+$
CCNN
CCNN$^+$

References

1. Hu, X., Mordohai, P.: A quantitative evaluation of confidence measures for stereo vision. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2012) 2121–2133
2. Merrell, P., Akbarzadeh, A., Wang, L., michael Frahm, J., Nistr, R.Y.D.: Real-time visibility-based fusion of depth maps. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2007)
3. Haeusler, R., Nair, R., Kondermann, D.: Ensemble learning for confidence measures in stereo vision. In: CVPR. Proceedings. (2013) 305–312 1.
4. Mordohai, P.: The self-aware matching measure for stereo. In: The International Conference on Computer Vision (ICCV), IEEE (2009) 1841–1848
5. Wedel, A., Meiner, A., Rabe, C., Franke, U., Cremers, D.: Detection and Segmentation of Independently Moving Objects from Dense Scene Flow. In: Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, Bonn, Germany, Springer (August 2009) 14–27
6. Di Stefano, L., Marchionni, M., Mattoccia, S.: A fast area-based stereo matching algorithm. Image and vision computing **22**(12) (2004) 983–1005
7. Spyropoulos, A., Komodakis, N., Mordohai, P.: Learning to detect ground control points for improving the accuracy of stereo matching. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2014) 1621–1628
8. Park, M.G., Yoon, K.J.: Leveraging stereo matching with learning-based confidence measures. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2015)
9. Poggi, M., Mattoccia, S.: Learning a general-purpose confidence measure based on o(1) features and a smarter aggregation strategy for semi global matching. In: Proceedings of the 4th International Conference on 3D Vision, 3DV. (2016)
10. Poggi, M., Mattoccia, S.: Learning from scratch a confidence measure. In: Proceedings of the 27th British Conference on Machine Vision, BMVC. (2016)