Leveraging Stereo Matching with Learning-based Confidence Measures (Supplementary Material)

Min-Gyu Park and Kuk-Jin Yoon
Computer Vision Laboratory, GIST, South Korea
{mpark,kjyoon}@gist.ac.kr

This supplementary material contains explanations of confidence measures and additional results for challenging datasets.

1. Confidence measures

We used a variety of confidence measures in [1–5] for constructing the 22-dimensional feature vector $\mathbf{f}^h$. Before explaning each of them, we introduce a few notations to explain confidence measures in a simple and clear way. Instead of using $C(\mathbf{p}, d)$ as a matching cost, we use c_i as the i^{th} minimum matching cost among all disparity hypotheses $\mathcal{D}$ for $\mathbf{p}$. d_i is the corresponding disparity value to c_i. For example, c_1 and d_1 are the minimum matching cost and the corresponding disparity value for $\mathbf{p}$. If necessary, we use $c_i(\mathbf{p})$ and $d_i(\mathbf{p})$ instead of c_i and d_i when the position $\mathbf{p}$ should be explicitly described. Hence, we also use a superscipt R to indicate that the matching cost or the disparity value is computed w.r.t. the right image. Otherwise, the matching cost is computed w.r.t. the left image which is the reference image. Lastly, we also define $\hat{c}_2$ as the second local minimum of the matching cost.

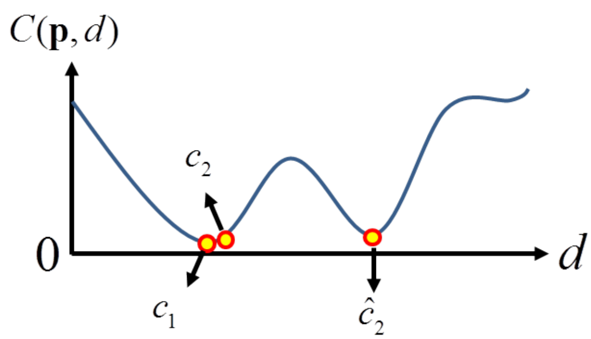


Figure 1. Matching costs with the minimum cost (c_1), the second minimum cost (c_2), and the second local minimum cost ($\hat{c}_2$)

1.1. Peak Ratio Measure (f_1)

The peak ratio [5] is defined as

$$f_1(\mathbf{p}) = \frac{\hat{c}_2}{c_1}, \tag{1}$$

which is the ratio of the second local minimum cost to the minimum matching cost. When the cost curve (matching costs) is unimodal, the second local minimum cannot be defined. Therefore, for some pixels, the peak ratio cannot be measured.

1.2. Naive Peak Ratio Measure (f_2)

The naive peak ratio [5] slightly changes the peak ratio as

$$f_2(\mathbf{p}) = \frac{c_2}{c_1}, \tag{2}$$

where the numberator is changed as the second minimum matching cost. In contrast to the peak ratio, the naive peak ratio can be computed for all pixels.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

1.3. Matching score measure (f_3)

The matching score measure [5] is defined as

$$f_3(\mathbf{p}) = -c_1, \tag{3}$$

which is the minimum matching cost itself. The less the matching score is, we believe that the match is more likely to be correct. Here, the minus sign is used to assign correct matches higher values.

1.4. Maximum margin measure (f_4)

The maximum margin [5] is defined as

$$f_4(\mathbf{p}) = c_2 - c_1, \tag{4}$$

which is the difference between the second and first minimum matching costs.

1.5. Winner margin measure (f_5)

Similary with the maximum margin, the winner margin measure [5] is defined as

$$f_5(\mathbf{p}) = \frac{c_2 - c_1}{\sum_i c_i}, \tag{5}$$

in which the denominator is used to normalize the maximum margin.

1.6. Maximum likelihood measure (f_6)

Confidence measures (or features) f_6 to f_9 treat a matching cost as a probability for a disparity value. The maximum likelihood measure [5] is defined as

$$f_6(\mathbf{p}) = \frac{\exp(-\frac{c_1}{2\sigma^2})}{\sum_i \exp(-\frac{c_i}{2\sigma^2})}, \tag{6}$$

where the denominator is the nomalization term, and the σ^2 is the variance of matching costs.

1.7. Perturbation measure (f_7)

The perturbation measure [4] is defined as

$$f_7(\mathbf{p}) = \sum_{i=2}^{|\mathcal{D}|} \exp(-\frac{(c_1 - c_i)}{\lambda^2}), \tag{7}$$

where λ is a scaling factor for the difference of matching costs. We set this as 1.

1.8. Negative entropy measure (f_8)

The negative entorpy measure [4, 5] is defined as

$$f_8(\mathbf{p}) = -\sum_i p(d_i) \log p(d_i), \tag{8}$$

where the probability density function $p(d)$ is computed as $p(d_i) = \frac{\exp(-c_i)}{\sum_k \exp(-c_k)}$.

1.9. Left-right difference measure (f_9)

The left-right difference measure [5] is defined as

$$f_9(\mathbf{p}) = \frac{c_2 - c_1}{|c_1 - c_1^R(\mathbf{p}_{d_1})|}, \tag{9}$$

where $c_1^R(\mathbf{p}_d)$ is the minimum matching cost of $\mathbf{p}_{d_1}$ w.r.t. the right image. $\mathbf{p}_{d_1}$ is the position $\mathbf{p}$ shifted by the d_1 along the scanline. Although d_1 is used for computing the confidence measure, we regard left-right difference is computed from matching costs.

1.10. Local curvature measure (f_{10})

The local curvature measure [4] is defined as

$$f_{10}(\mathbf{p}) = -2C(\mathbf{p}, d_1) + C(\mathbf{p}, d_1 - 1) + C(\mathbf{p}, d_1 + 1). \tag{10}$$

1.11. Disparity variance measure (f_{11} to f_{14})

The disparity variance measure [4] is defined as

$$f_{11}(\mathbf{p}) = -\frac{1}{|N_{\mathbf{p}}| - 1} \sum_{\mathbf{p} \in N_{\mathbf{p}}} (d_1(\mathbf{q}) - \mu)^2, \tag{11}$$

where $N_{\mathbf{p}}$ is a set of neighboring pixels for $\mathbf{p}$, and the minus sign is used to assign correct matches higher values. In fact, the disparity variance measure in [4] used the magnitude of disparity gradients instead of computing the variance of disparity values.

1.12. Distance from discontinuity (f_{15})

The distance from discontinuity measure [1] is defined as

$$f_{15}(\mathbf{p}) = \mathrm{dist}(\mathbf{p}, \mathbf{q}), \tag{12}$$

where $\mathbf{q}$ is the position of the nearest discontinuity from $\mathbf{p}$. In [1], a pixel is determined as the (potential) depth discontinuity if $d_1(\mathbf{q})$ is different from its four neighbors. However, we determined the depth discontinuity in a slightly different way. We computed the disparity gradient, and deteremined as the (potential) depth discontinuity if the magnitude of gradient is larger than 2.

1.13. Median deviation measure (f_{16} to f_{19})

The median deviation measure [1] is defines as

$$f_{16}(\mathbf{p}) = -|d_1 - \underset{\mathbf{q} \in N_{\mathbf{p}}}{\mathrm{MED}}(d_1(\mathbf{q}))|, \tag{13}$$

which is the difference between d_1 and the median of disparity values. In [1], the difference is truncated at 2, but we did not truncate the difference value. Moreover, [1] computed the median value in a 5×5 window; but, we used various window sizes, 5×5 for f_{16}, 7×7 for f_{17}, 9×9 for f_{18}, and 11×11 for f_{19}.

1.14. Left-right consistency (f_{20})

The left-right consistency measure [5] is defined as

$$f_{20}(\mathbf{p}) = -|d_1 - d_1^R(\mathbf{p}_{d_1})|, \tag{14}$$

which is the difference between the disparity value and its corresponding disparity value in the right image.

1.15. Magnitute of image gradient (f_{21})

The image gradient measure [4] is defined as

$$f_{21}(\mathbf{p}) = ||\nabla I(\mathbf{p})||, \tag{15}$$

which is the magnitude of the image gradient at $\mathbf{p}$.

1.16. Distance from border (f_{22})

Lastly, the distance from border measure [1] is defined as

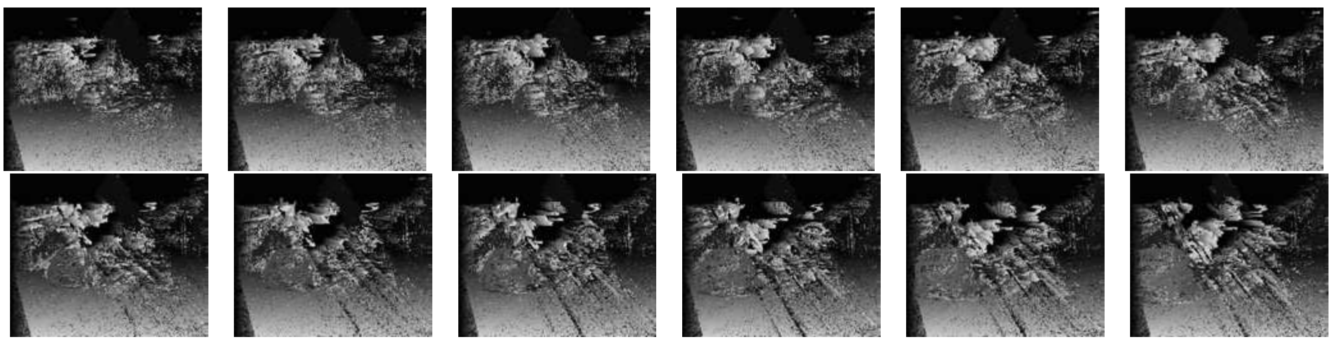
$$f_{22}(\mathbf{p}) = -\min(p_x, d_{\max}), \tag{16}$$

where p_x is the position of $\mathbf{p}$ in the x-axis, and $d_{\max}$ is the maximum disparity value. In [1], this measure is defined as the distance from the image border. Instead, we only considered the left most image border (when the left image is the reference) in which pixels do not have actual correspondences due to the different field of views between the left and right cameras.

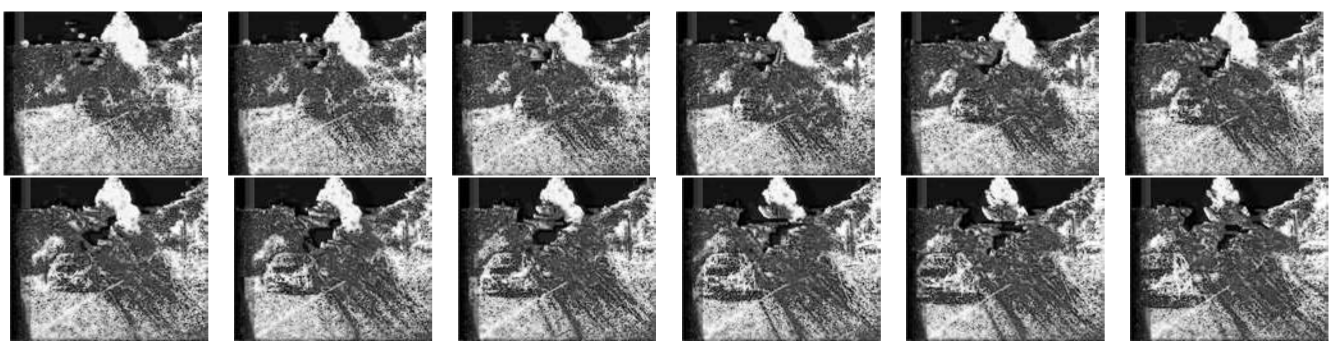
2. Qualitative evaluation

In this section, we show additional results for challenging datasets [6].

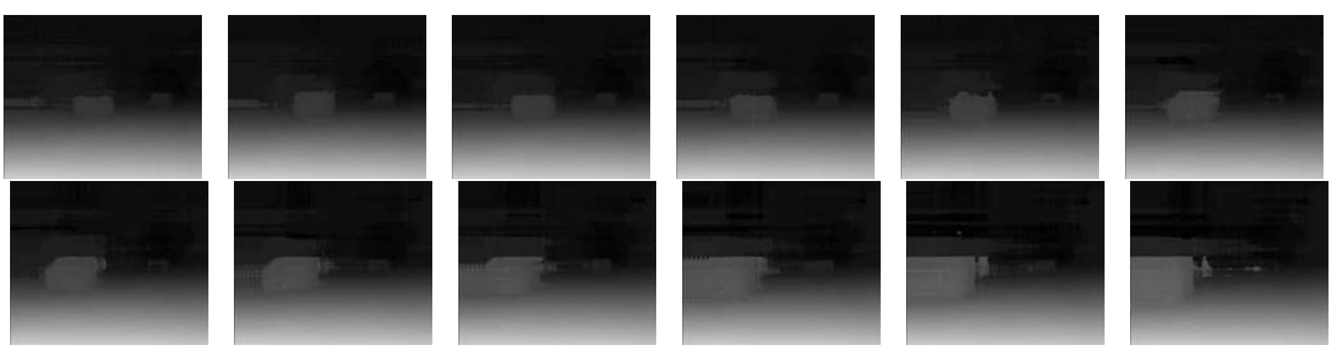
(a) Left images



(b) Initial disparity maps



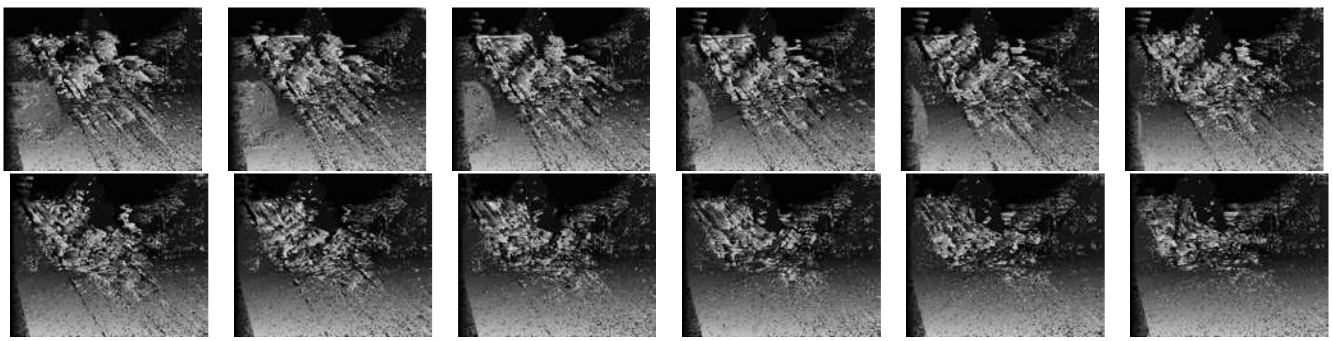
(c) Predicted confidence maps



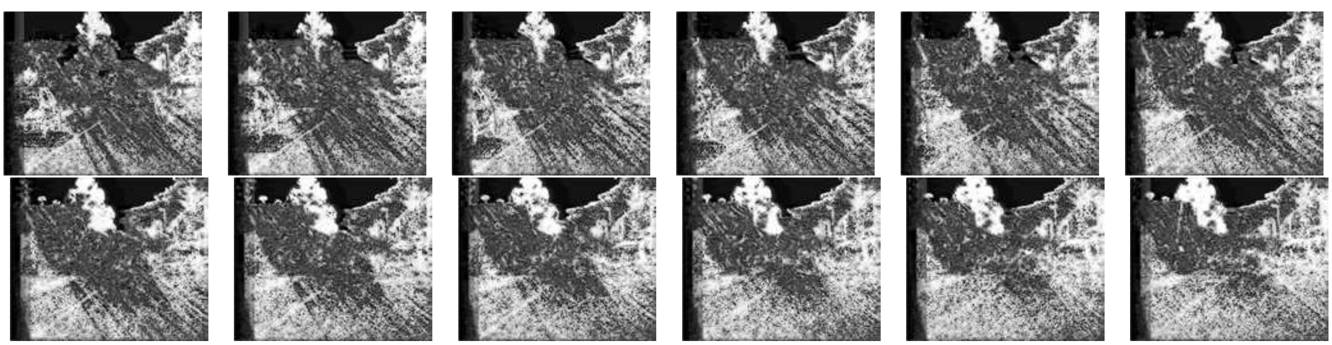
(d) Estimated disparity maps

Figure 2. Results for 7-18 frames (sun flare sequence)

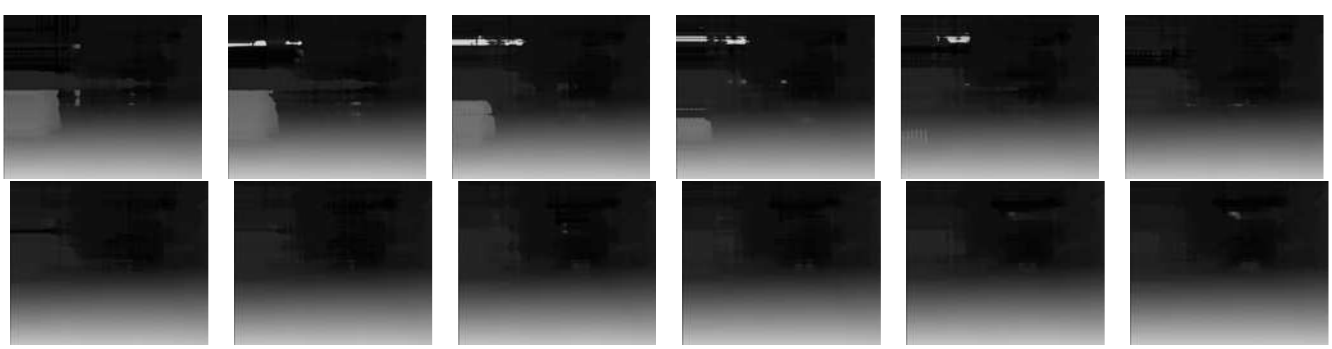
(a) Left images



(b) Initial disparity maps



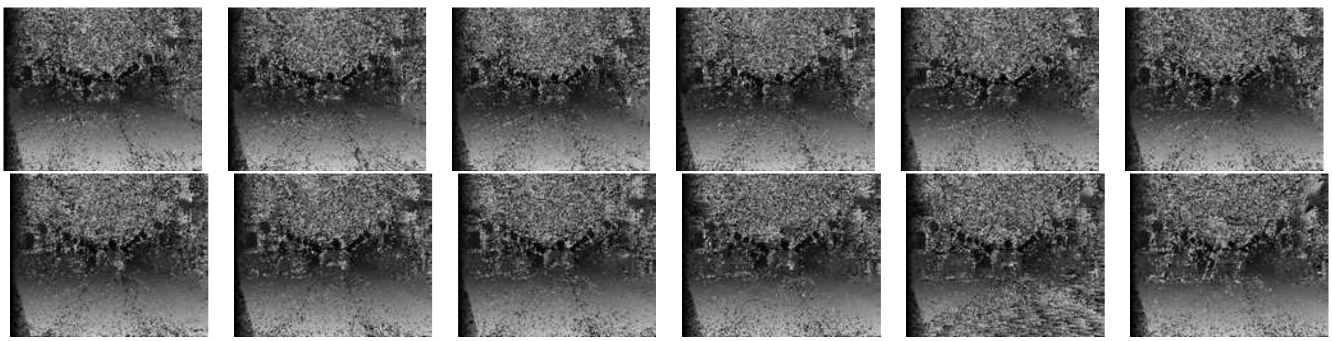
(c) Predicted confidence maps



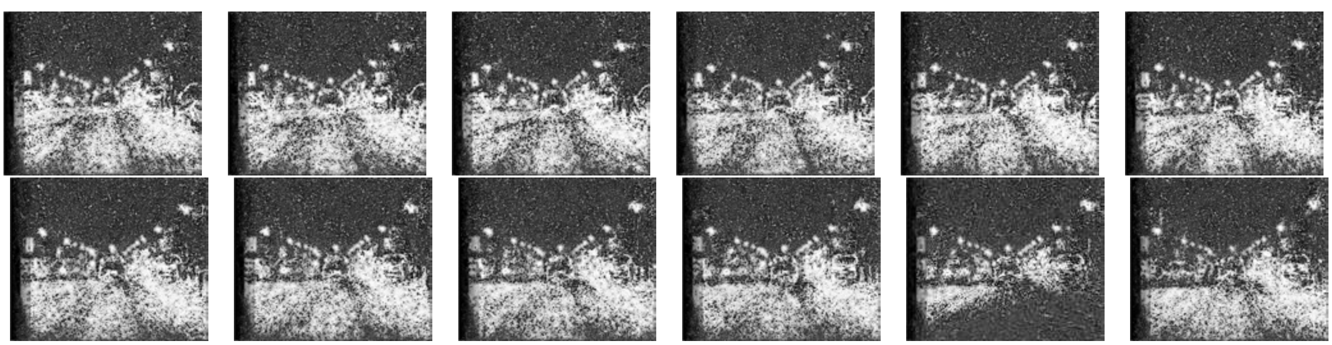
(d) Estimated disparity maps

Figure 3. Results for 19-30 frames (sun flare sequence)

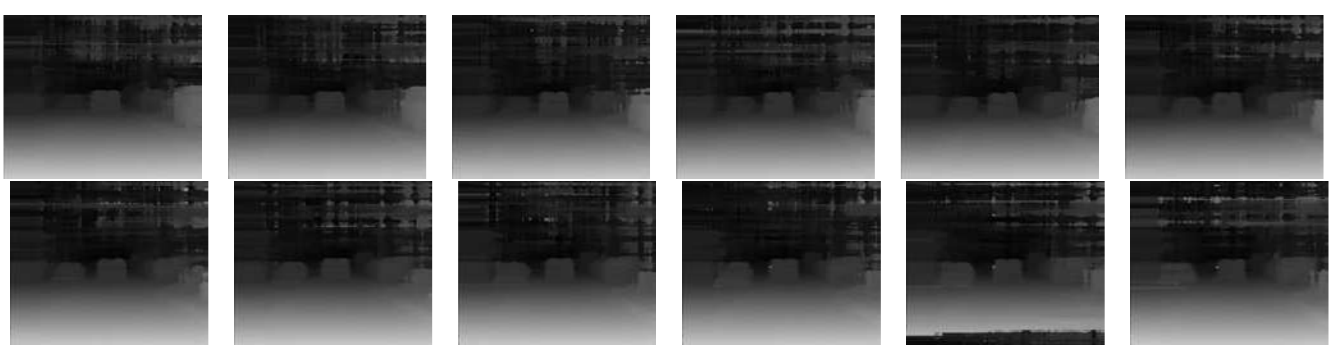
(a) Left images



(b) Initial disparity maps



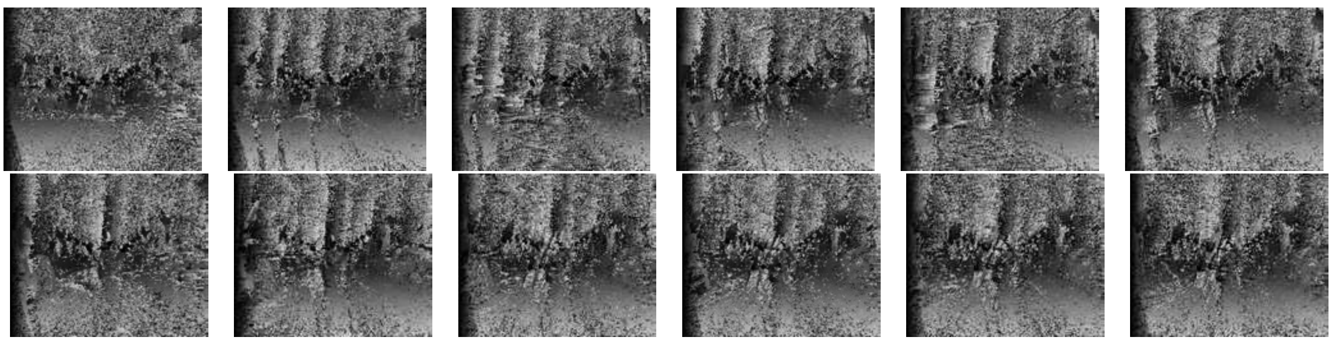
(c) Predicted confidence maps



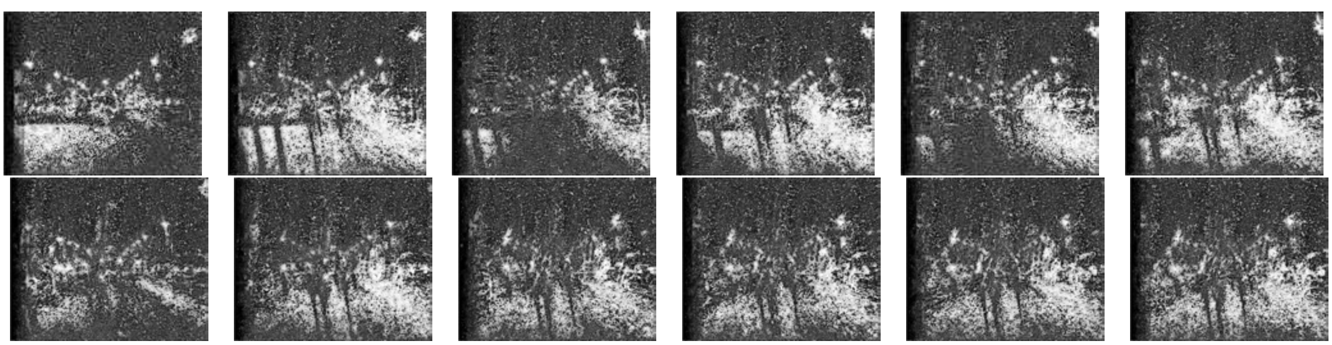
(d) Estimated disparity maps

Figure 4. Results for 7-18 frames (night and snow sequence)

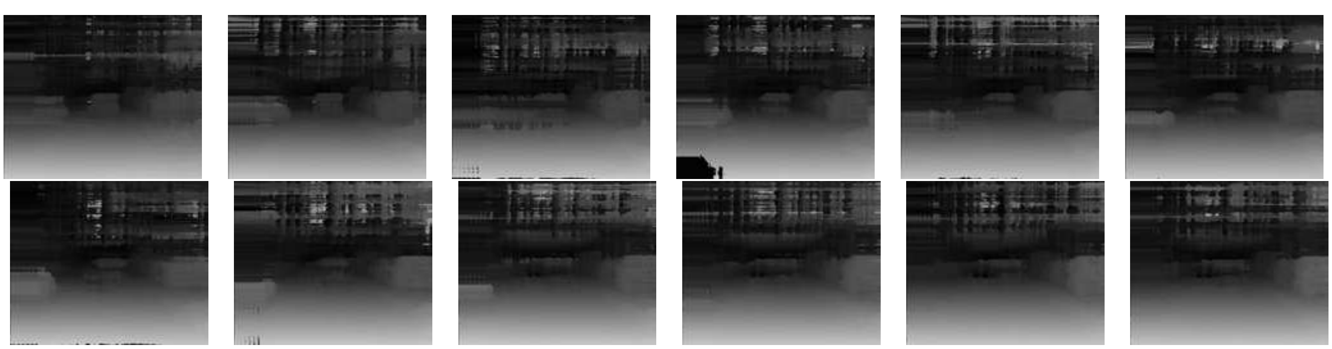
(a) Left images



(b) Initial disparity maps

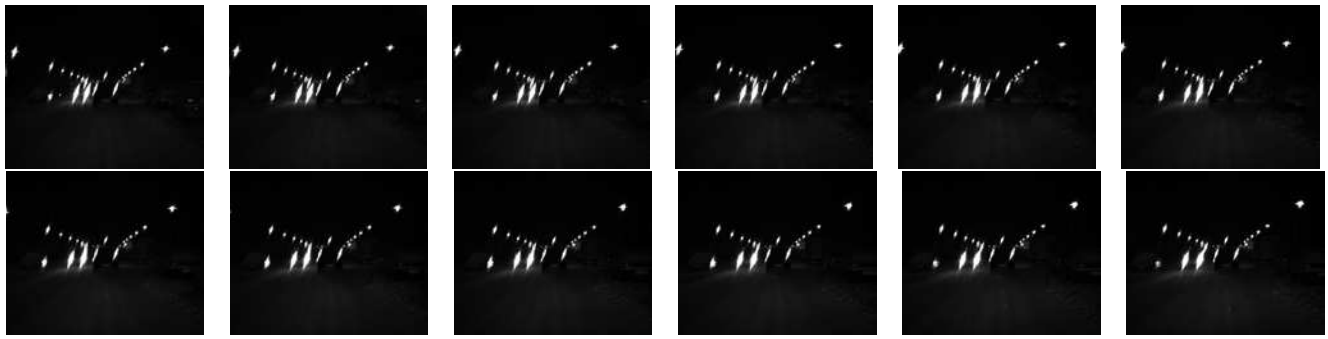


(c) Predicted confidence maps

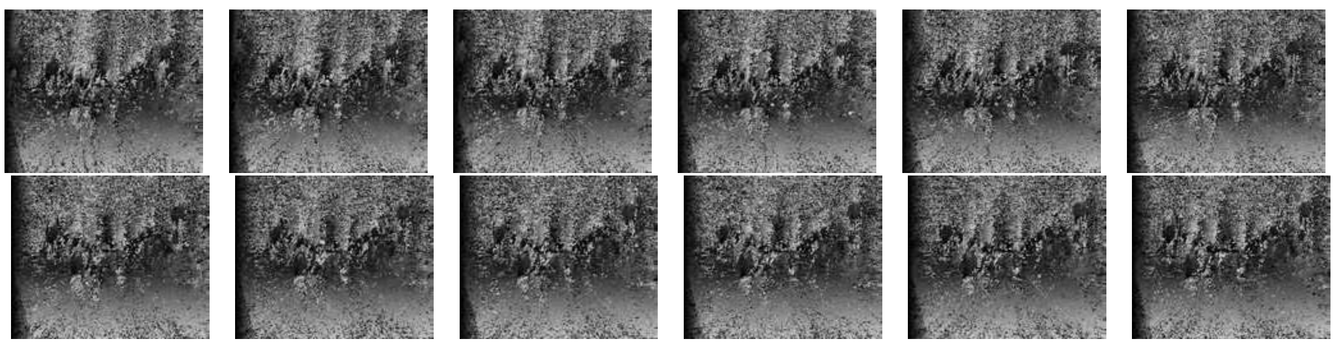


(d) Estimated disparity maps

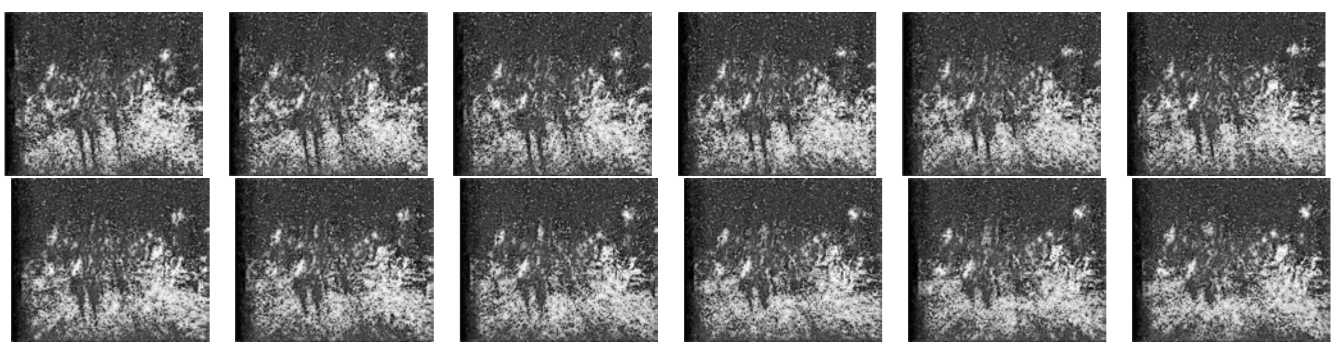
Figure 5. Results for 19-30 frames (night and snow sequence)



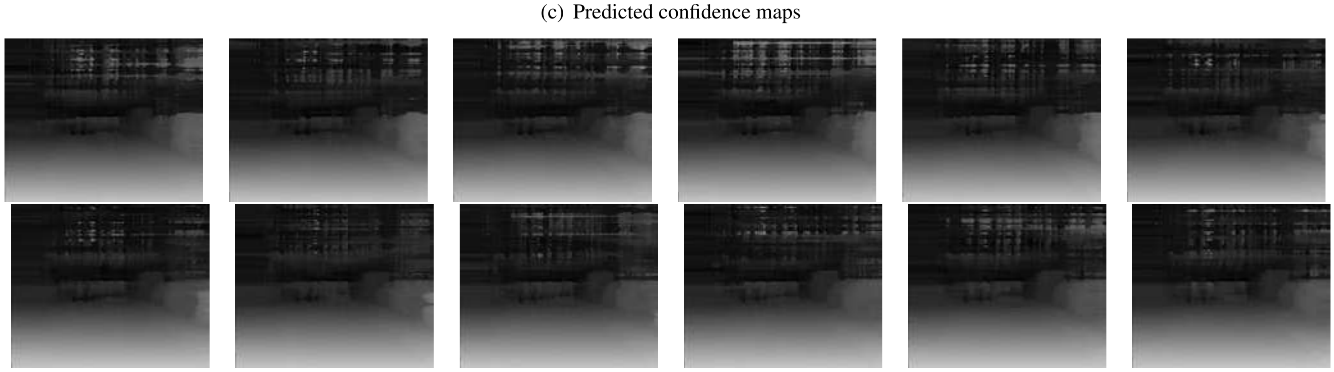
(a) Left images



(b) Initial disparity maps



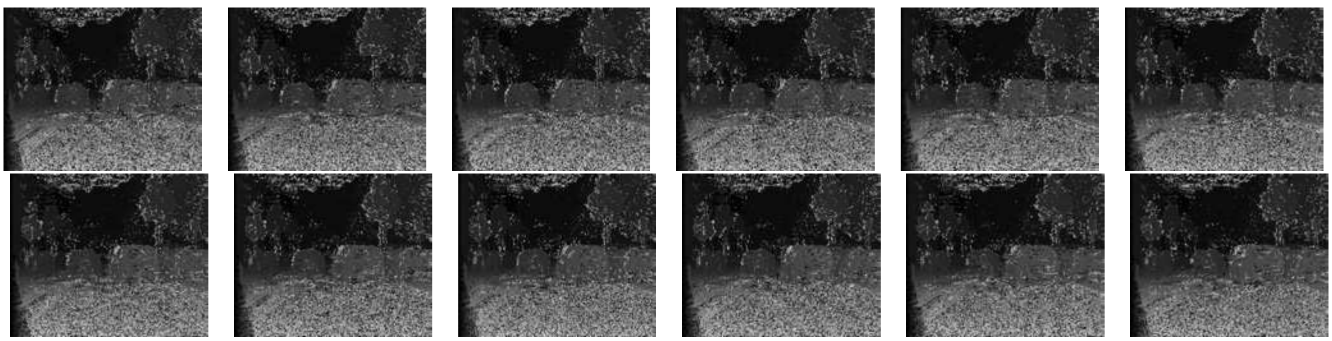
(c) Predicted confidence maps



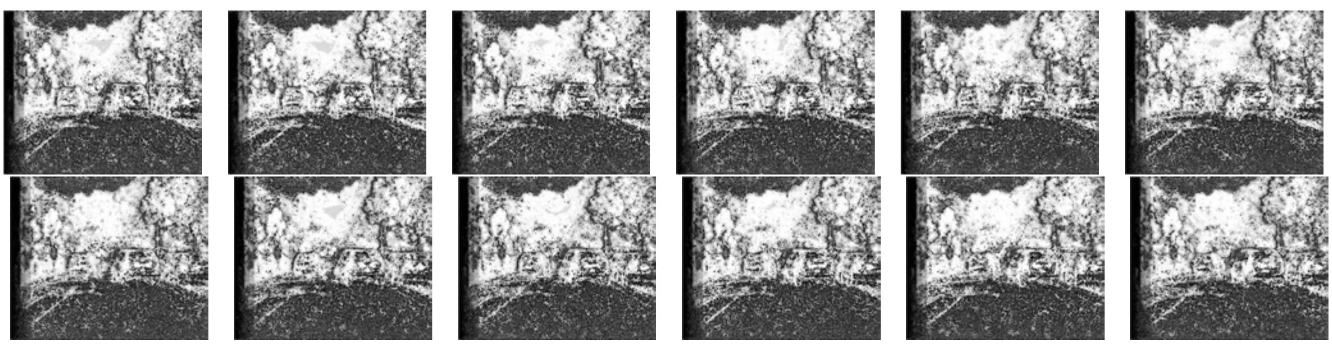
(d) Estimated disparity maps

Figure 6. Results for 31-42 frames (night and snow sequence)

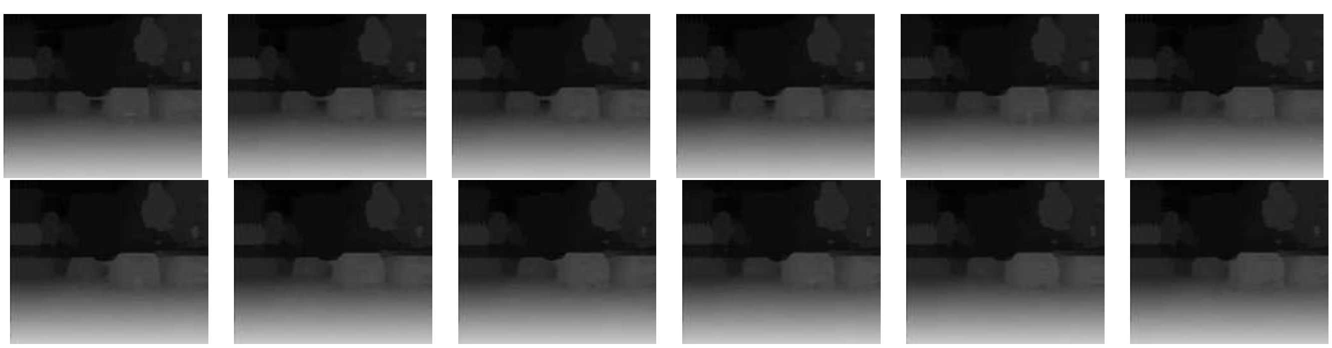
(a) Left images



(b) Initial disparity maps



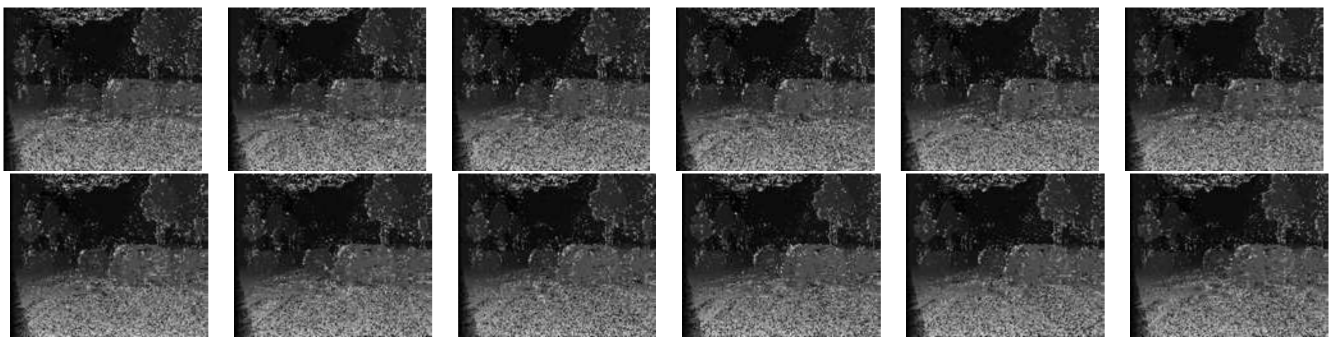
(c) Predicted confidence maps



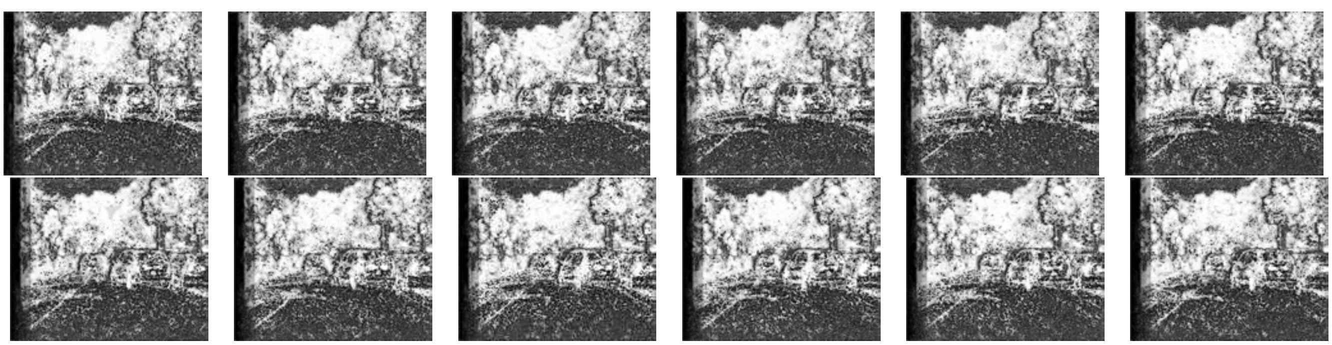
(d) Estimated disparity maps

Figure 7. Results for 7-18 frames (reflecting car sequence)

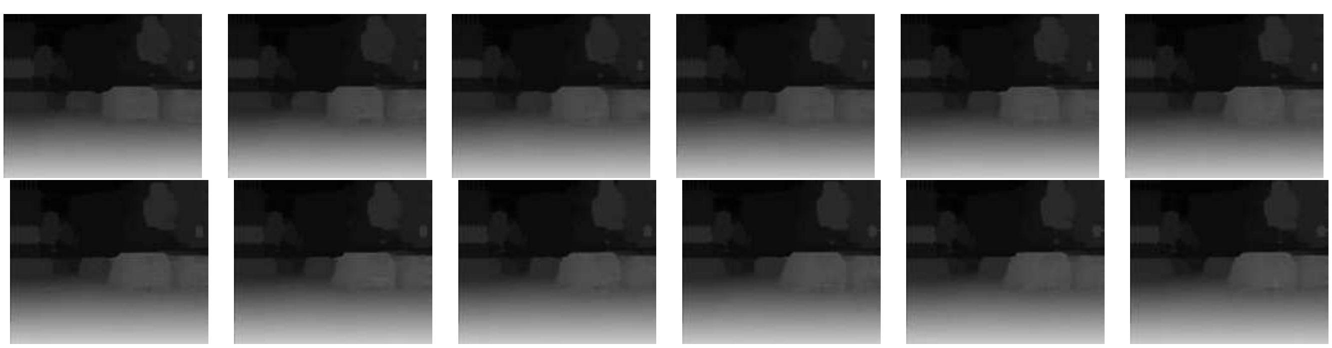
(a) Left images



(b) Initial disparity maps



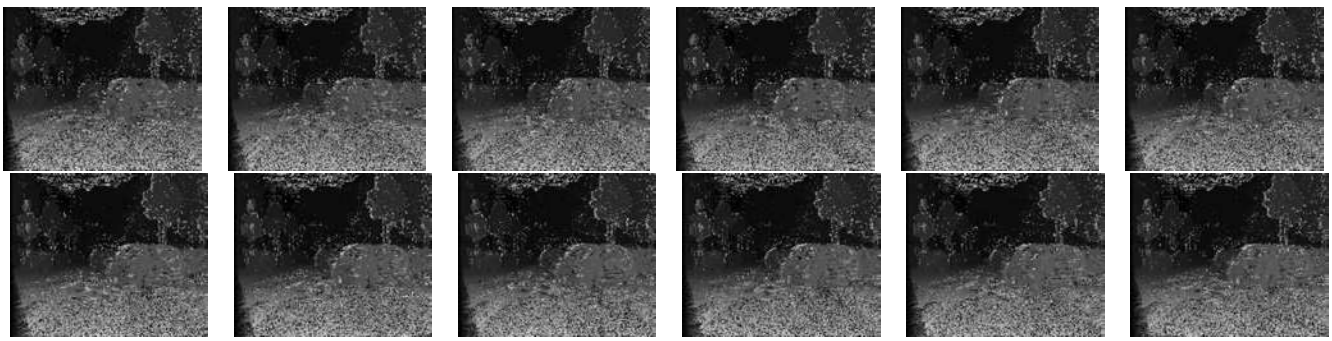
(c) Predicted confidence maps



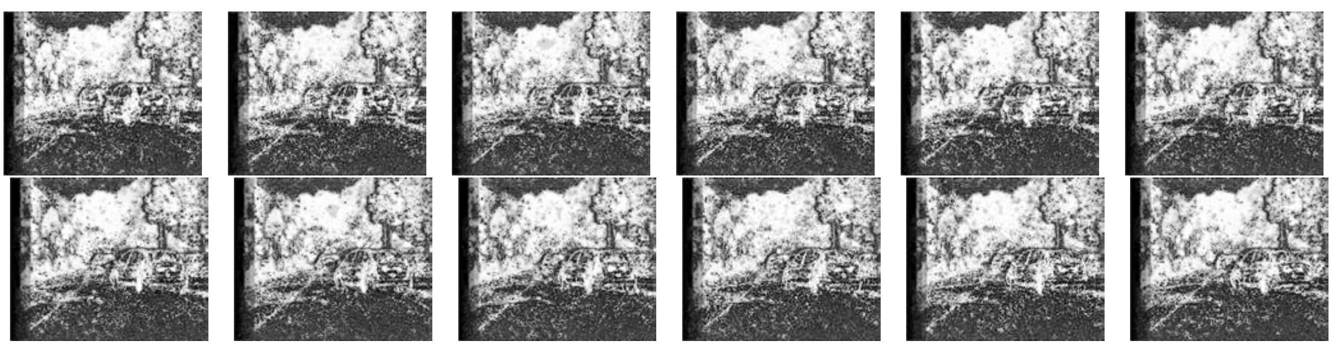
(d) Estimated disparity maps

Figure 8. Results for 19-30 frames (reflecting car sequence)

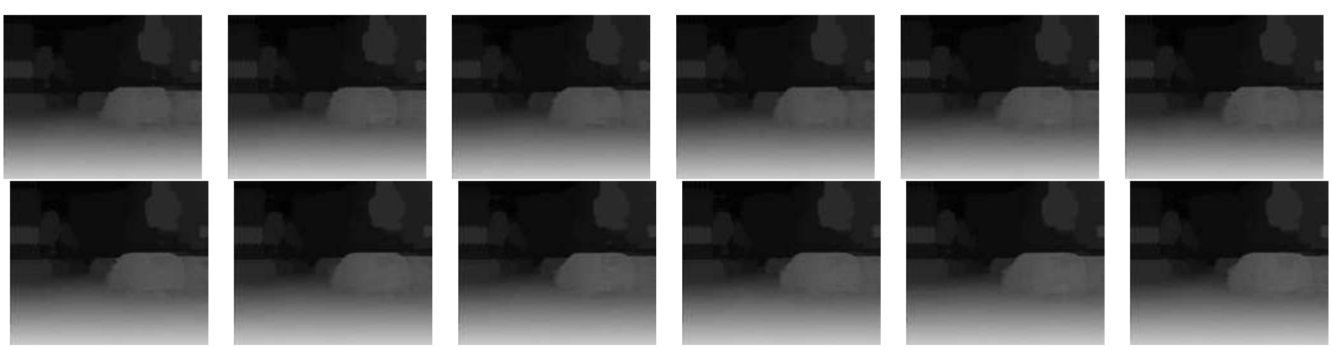
(a) Left images



(b) Initial disparity maps



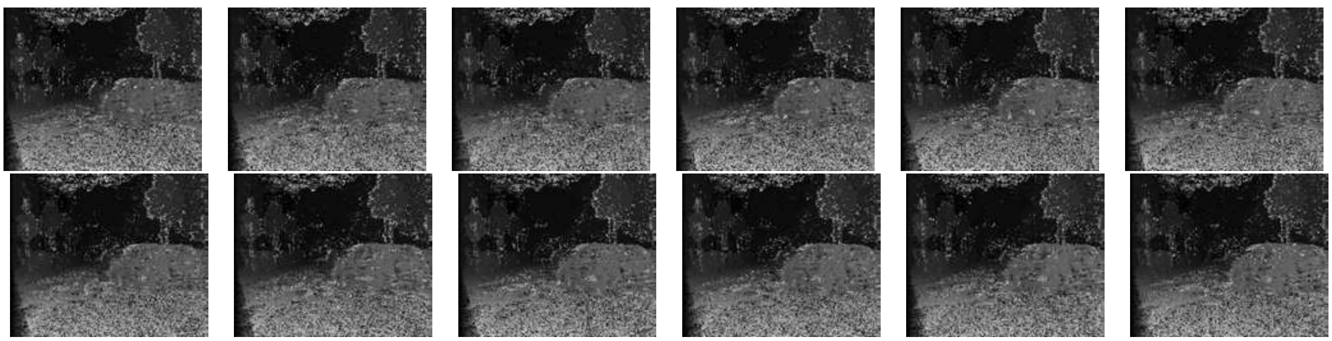
(c) Predicted confidence maps



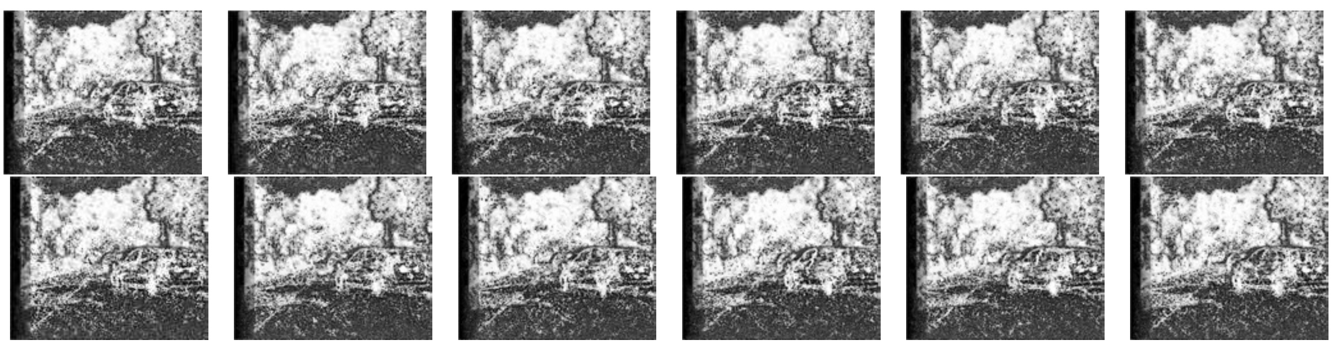
(d) Estimated disparity maps

Figure 9. Results for 31-42 frames (reflecting car sequence)

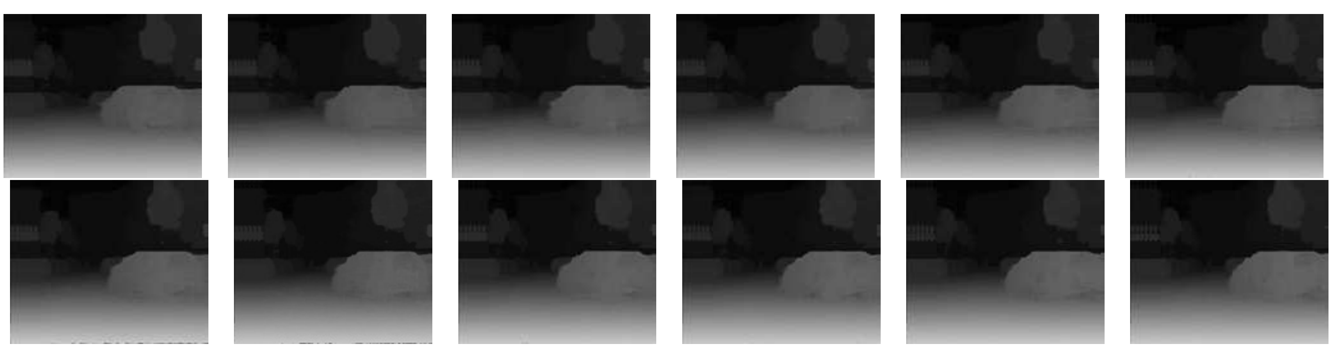
(a) Left images



(b) Initial disparity maps



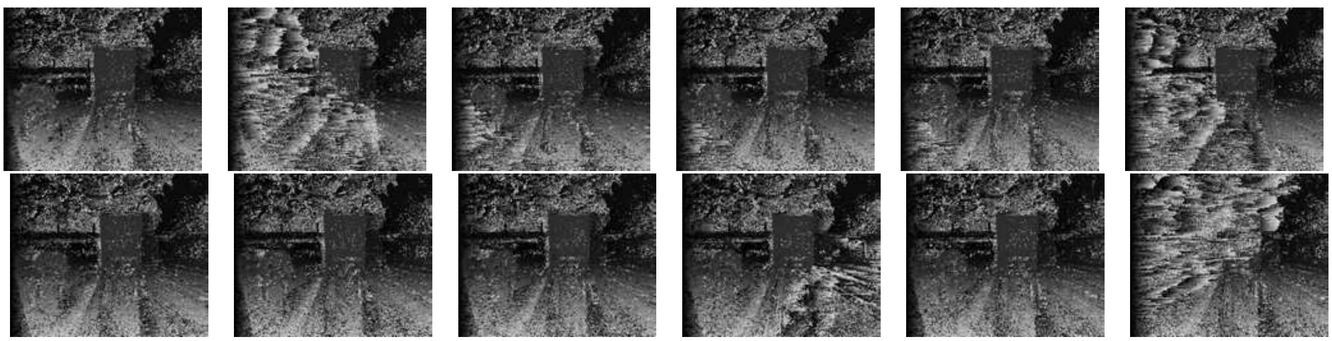
(c) Predicted confidence maps



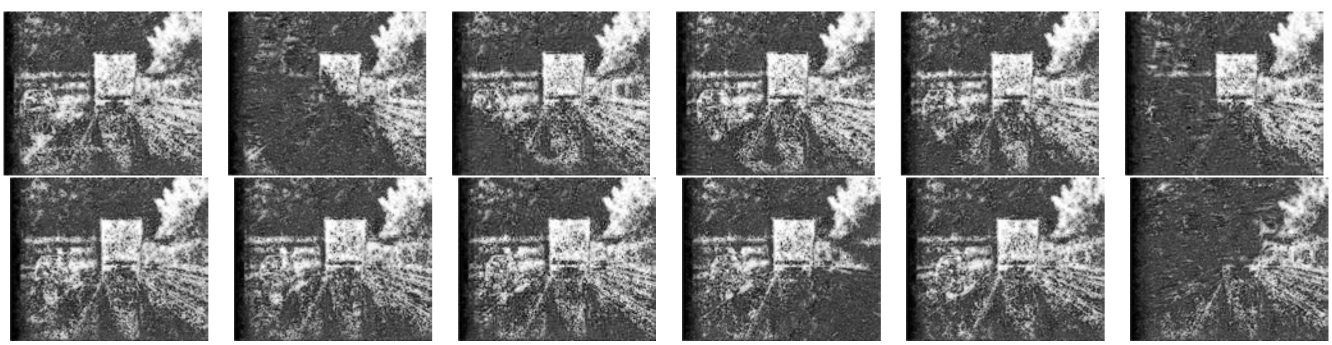
(d) Estimated disparity maps

Figure 10. Results for 43-54 frames (reflecting car sequence)

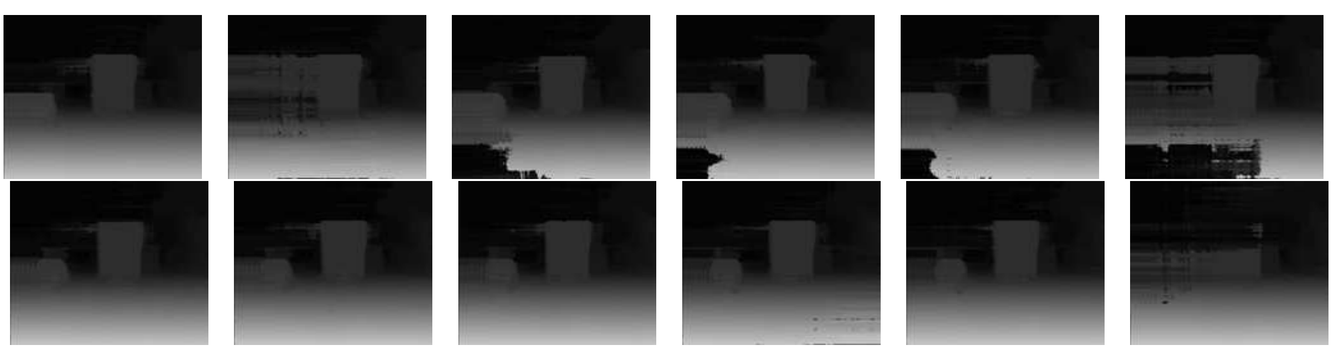
(a) Left images



(b) Initial disparity maps



(c) Predicted confidence maps



(d) Estimated disparity maps

Figure 11. Results for 7-18 frames (wet autobahn sequence)

References

[1] N. K. Aristotle Spyropoulos and P. Mordohai. Learning to detect ground control points for improving the accuracy of stereo matching. In *CVPR*, 2014.

[2] G. Egnal, M. Mintz, and R. P. Wildes. A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and Vision Computing*, 22(12):943 – 957, 2004.

[3] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *PAMI*, 24(8):1127–1133, 2002.

[4] R. Haeusler, R. Nair, and D. Kondermann. Ensemble learning for confidence measures in stereo vision. In *CVPR*, pages 305–312, 2013. 1.

[5] X. Hu and P. Mordohai. A quantitative evaluation of confidence measures for stereo vision. *PAMI*, 34(11):2121–2133, 2012.

[6] S. Meister, B. Jähne, and D. Kondermann. Outdoor stereo camera system for the generation of real-world benchmark data sets. *Optical Engineering*, 51(02):021107, 2012.