over the most recent frames. By maintaining a reasonably high value of γ_t, the auxiliary cost also preserves temporal edges, essentially reducing over-smoothing of a pixel's disparity when a pixel transitions from one depth to another in subsequent frames.

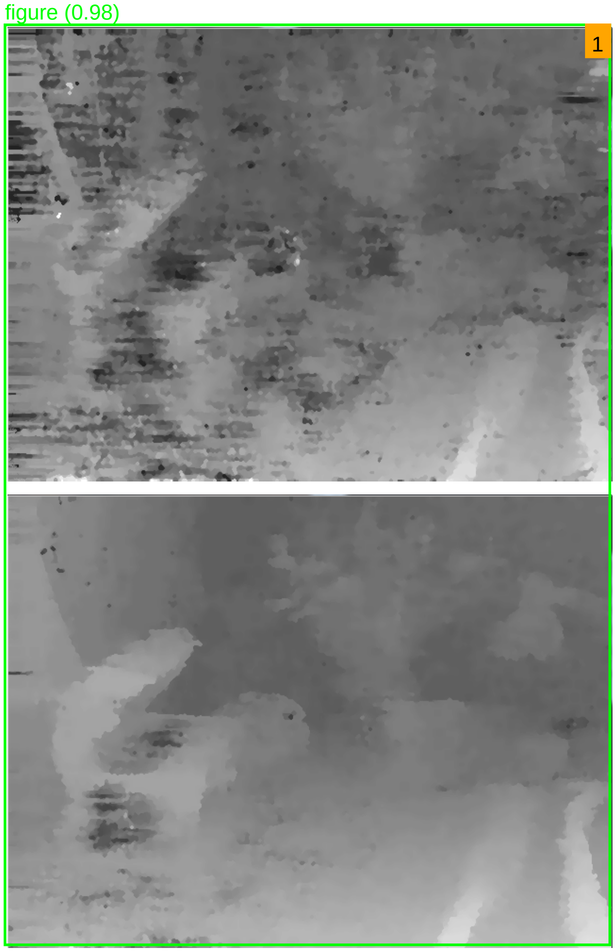

Figure 5: A comparison of stereo matching without temporal cost aggregation (top) and with temporal cost aggregation (bottom) for a single frame in the synthetic video sequence where the noise is ±30 and the feedback coefficient is $\lambda = 0.8$.

The optimal value of the feedback coefficient is largely dependent on the noise being added to the image. Figure 4 shows the optimal values of λ for noise ranging between ±0 to ±40. As intuition would suggest, it is more beneficial to rely on the auxiliary cost when noise is high and it is more beneficial to rely on the current cost when noise is low. Figure 5 illustrates the improvements that are achieved when applying temporal stereo matching to a particular pair of frames in the synthetic video sequence. Clearly, the noise in the disparity map is drastically reduced when temporal stereo matching is used.

The algorithm was implement using NVIDIA's Compute Unified Device Architecture (CUDA). The details of the implementation are similar to those given in [3]. When compared to other existing real-time stereo matching implementations, the proposed implementation achieves the highest speed of operation measured by the number of disparity hypotheses evaluated per second, as shown in Table II. It is also the second most accurate real-time method in terms of error rate, as measured using the Middlebury stereo evaluation benchmark. It should be noted that it is difficult to establish an unbiased metric for speed comparisons, as the architecture, number of cores, and clock speed of graphics hardware used are not consistent across implementations.

Table II: A comparison of speed and accuracy for the implementations of many leading real-time stereo matching methods.

Method	GPU	MDE/s[1]	FPS[2]	Error[3]
Our Method	GeForce GTX 680	215.7	90	6.20
CostFilter [10]	GeForce GTX 480	57.9	24	5.55
FastBilateral [7]	Tesla C2070	50.6	21	7.31
RealtimeBFV [8]	GeForce 8800 GTX	114.3	46	7.65
RealtimeBP [21]	GeForce 7900 GTX	20.9	8	7.69
ESAW [6]	GeForce 8800 GTX	194.8	79	8.21
RealTimeGPU [5]	Radeon XL1800	52.8	21	9.82
DCBGrid [19]	Quadro FX 5800	25.1	10	10.90

[1] Millions of Disparity Estimates per Second.
[2] Assumes 320×240 images with 32 disparity levels.
[3] As measured by the Middlebury stereo performance benchmark using the avgerage % of bad pixels.

V. Conclusion

While the majority of stereo matching algorithms focus on achieving high accuracy on still images, the volume of research aimed at recovery of temporally consistent disparity maps remains disproportionally small. This paper introduces an efficient temporal cost aggregation scheme that can easily be combined with conventional spatial cost aggregation to improve the accuracy of stereo matching when operating on video sequences. A synthetic video sequence, along with ground truth disparity data, was generated to evaluate the performance of the proposed method. It was shown that temporal aggregation is significantly more robust to noise than a method that only considers the current stereo frames.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, April-June 2002.

[2] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 195–202, June 2003.

[3] J. Kowalczuk, E. Psota, and L. Perez, "Real-time stereo matching on CUDA using an iterative refinement method for adaptive support-weight correspondences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 23, pp. 94 –104, Jan. 2013.

[4] K.-J. Yoon and I.-S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, (Washington, DC, USA), pp. 924–931, IEEE Computer Society, 2005.

[5] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nister, "High-quality real-time stereo using adaptive cost aggregation and dynamic programming," in *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, (Washington, DC, USA), pp. 798–805, IEEE Computer Society, 2006.