Edge aware disparity estimation for intermediate view synthesis

M. Farre*, H. Lakshman*, P. Helle*, H. Schwarz*, K. Mueller*, D. Marpe* and T. Wiegand*‡

*Image Processing Department, Fraunhofer Institute for Telecommunications – Heinrich Hertz Institute, Berlin, Germany

‡Image Communication Chair, Technical University of Berlin, Germany

***Abstract*— In this paper we present a new local and multiscale disparity estimation algorithm. Results show that the proposed method can preserve arbitrarily shaped depth discontinuities. Some results obtained from this method are shown as depth maps and synthetic views produced using the estimated disparity maps.**

I. INTRODUCTION AND RELATED WORK

The disparity estimation problem has led researchers to develop a wide range of techniques, which can be generally classified into local and global methods. Local approaches make use of color or intensity values within a finite window to determine the disparity value for each pixel. Global approaches, on the other hand, formulate the stereo matching problem as an energy function and solve it using various optimization methods such as belief propagation [1], graph cuts [2], etc. A comprehensive evaluation of dense stereo matching algorithms for rectified image pairs can be found in [3].

When the local structures of the image pixels are similar, finding their correspondences in other images without global perception can be very challenging. To properly deal with these ambiguity problems, local methods generally obtain support from neighboring pixels. Several of the proposed techniques differ in the way they aggregate pixels for support purposes, some of them are: multiple window methods [4], adaptive shape methods [5] or adaptive weight methods [6] which according to the evaluation in [7] outperform approaches that explicitly modify the shape of their supports. For more details, a classification and evaluation of these different techniques for support aggregation can be found in [7].

In adaptive weight methods, each pixel located in the support region has a weight, which is updated according to its intensity difference and spatial distance to the pixel under study. The general idea is that pixels with similar intensity and small distance to the pixel under study are more likely to have the same depth as the pixel being analyzed. Among the large number of applications of the bilateral filter [8], some researchers have used its weight or an approximation (because it is computationally expensive [7]) as adaptive weights for disparity [9] or optical flow [10] estimation. This provides in some cases, estimations comparable to the ones obtained with global techniques.

In this paper we present a local and multiscale disparity estimation algorithm that makes an inexpensive computational use of the bilateral filter. Making use of the datasets [11] and renderer mentioned in the 3D High Efficiency Video Coding (HEVC) test model [12] we compare our estimated depth maps with the ones provided in the datasets and we use them to render synthetic views, showing rendering improvement near arbitrarily shaped depth discontinuities.

II. PROPOSED ALGORITHM

An overview of the proposed disparity estimation algorithm is shown in Fig. 1, which iterates a multiscale approach assumimg rectified stereo views as input. The disparity is calculated starting from the coarsest level of an L level Gaussian Pyramid and moving to the finer levels. The resulting disparities of one level are used as the input to the next finer one. Our algorithm pipeline is an adaptation of the one described in [10] for optical flow estimation, maintaining its multiscale approach and its idea of candidate cost averaging as smoothing prior. It is extended by adding new measures for consistency check and refinement to be applied in a disparity estimation context.

For each level, the algorithm pipeline operates in two parallel paths. One path uses the left view as the reference and the right view as the target, while the other path uses them in the opposite way. In each of these paths, a cost vector is generated and the average cost is minimized. Both paths join again for a disparity consistency check and finally a refinement of the disparity values is performed. The result is either upsampled and used as input into the immediately finer pyramid level or is presented as the final disparity map in case of the last iteration. In the next sub-sections the different stages that form the algorithm are described.

A. Calculation and evaluation of the disparity candidates

For each pixel p in the reference image I_r we search for its disparity value on the current level d_l contemplating as new disparity candidates $\hat{d}$ a range R of values on both sides of the current disparity value, which is 0 in the coarsest level of the pyramid or the upsampled value of the next coarser level d'_{l+1} of the pyramid otherwise. A matching cost vector C is obtained with the matching cost of the different candidates. This matching cost corresponds to the color difference between p in I_r and the pixel pointed by $\hat{d}$ in the target view I_t

$$C(\hat{d}_i) = \left\| I_r(p) - I_t(p + \hat{d}_i) \right\|^2. \quad (1)$$

A first estimation evaluating C is obtained, which helps to classify the trustworthiness of the candidates:

$$W_t = mean(C) - \min(C). \quad (2)$$

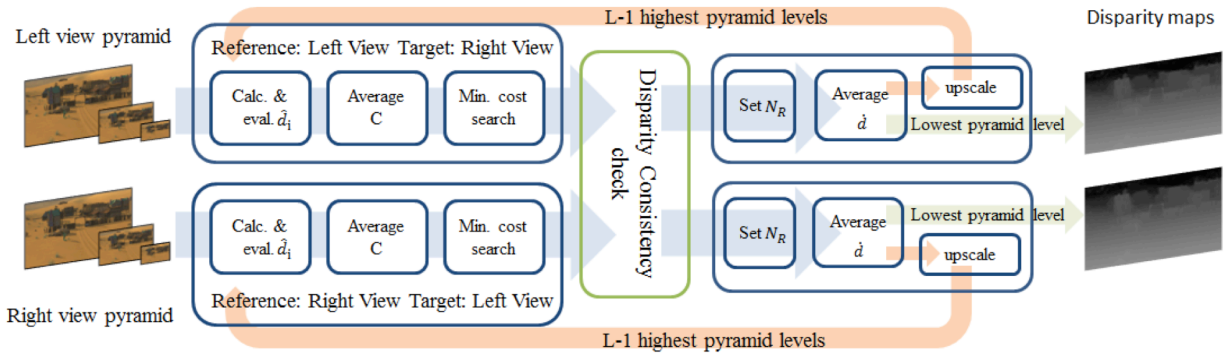


Figure 1: Diagram with the different steps involved in the presented algorithm.

For instance, a small W_t value indicates that all the disparity candidates in C are very similar, which does not help to decide for one of them.

B. *Averaging of the cost matching vector*

Depth is assumed to be smooth in a neighborhood N_{avg} in order to avoid the addition of terms that link the disparity estimation between different pixels, making the optimization problem harder to solve. We assume that a correct disparity value for a given pixel has to be a good representation of the disparity for the pixels in N_{avg} .

For each disparity candidate $\hat{d}_i$ of the pixel under analysis p_0, we search for the disparity candidates $\hat{d}_j$ from the other pixels p_k within N_{avg} that fulfill $\hat{d}_i(p_0) = \hat{d}_j(p_k)$ and we group these disparity candidates with their respective matching cost $C(\hat{d}_j)$ information and pixel p_k color and position information in the subset Ω. The average matching cost C' for every candidate $\hat{d}_i(p_0)$, is calculated as:

$$C'(\hat{d}_i(p_0)) = \frac{\sum_\Omega W_t(p_k) W_{BIL}(p_0,p_k) C(\hat{d}_j(p_k))}{\sum_\Omega W_t(p_k) W_{BIL}(p_0,p_k)} \quad (3)$$

where $W_{BIL}(p_0, p_k)$ is the bilateral filter weight between the pixels p_0 and p_k, and $X_0$and X_k are the positions of the pixels $p_0$and p_k respectively, so that,

$$W_{BIL}(p_0, p_k) = e^{-\frac{1}{2}\left(\frac{X_0 - X_k}{\sigma_x}\right)^2} e^{-\frac{1}{2}\left(\frac{I(P_0) - I(P_k)}{\sigma_c}\right)^2} \quad (4)$$

C. *Minimum cost search*

After the cost averaging, a minimum cost search using a Winner-Take-All (WTA) criteria is performed, $\dot{d}$ being the final pre-refinement disparity value $\dot{d} = argmin_{\hat{d}_i}(C'(\hat{d}_i))$.

D. *Disparity Consistency check*

Pre-refinement disparity values $\dot{d}$ for all the pixels are calculated and grouped in maps: one map with the left to right disparity estimations D_{LR} and the other one with the right to left estimations D_{RL}. For every disparity map, a mask M will be filled, setting its values to 1 if the disparity of the pixel under study is consistent and a 0 if it is not. Three consistency checks are now defined: the first one has to be passed in all the pyramid levels and the second and third one will be considered only in the finest level of the pyramid under specific conditions.

1) *Left to Right and Right to Left consistency check*

Below, the rule to fill M is described for the case of the disparity left to right M_{LR}.

$$M_{LR}(p) = \begin{cases} 1 & D_{LR}(p) + D_{RL}\big(p + D_{LR}(p)\big) < \tau \\ 0 & D_{LR}(p) + D_{RL}\big(p + D_{LR}(p)\big) > \tau \end{cases} \quad (5)$$

The other case M_{RL} can be deduced from it. τ is set empirically.

2) *Provided range of possible disparity values*

If the range of possible disparity values for our problem is known, any disparity value outside the range is marked as inconsistent in M.

3) *Stereo baseline with parallel image planes*

If the footage is obtained from a stereo baseline with parallel image planes configuration, only negative pixel position offsets are expected in the D_{LR} estimation and only positive values in the D_{RL} estimation. Disparity values not fulfilling this rule can be marked as inconsistent in their respective M.

E. *Refinement and upsampling*

Disparity values are averaged within their neighborhood N_R using weights from the bilateral filter and without taking into account disparities corresponding to the pixels that did not pass the consistency check.

$$d(p_0) = \frac{\sum_{Pi \in N_R} M(p_i) W_{BIL}(p_0,p_i) \dot{d}(p_i)}{\sum_{Pi \in N_R} M(p_i) W_{BIL}(p_0,p_i)} \quad (6)$$

The size of N_R in (6) is empirically fixed but it can grow if only inconsistent disparity values are found. The new neighborhood size N'_R can grow until the following expression is fulfilled

$$\frac{\sum_{N'_R} M(p_i)}{N_R} > \beta. \quad (7)$$

The improvements achieved with the N_R growing constraint can be seen in Fig. 2.

1) L-1 first levels

The disparity map is upsampled spatially by means of sample replication, disparity values are scaled by a factor of two. The upsampled disparity map will be used as input for the next pyramid level.

2) Last level

Equation (6) is modified by adding a new weight W_D responsible for evaluating the disparity similarity,

$$W_D(p_0, p_i) = e^{-\frac{1}{2}\left(\frac{\dot{d}(P_0)-\dot{d}(P_i)}{\sigma_d}\right)^2}. \quad (8)$$

It minimizes the blending of disparities from neighbor pixels with similar colors but with substantially different depths. Adding the new weight (8) to (6) we obtain

$$d_l(p_0) = \frac{\sum_{Pi \in N_{R'}} M(p_i) W_{BIL}(p_0,p_i) W_D(p_0,p_i) \dot{d}(p_i)}{\sum_{Pi \in N_{R'}} M(p_i) W_{BIL}(p_0,p_i) W_D(p_0,p_i)}. \quad (9)$$

Its impact in the disparity estimation is shown in Fig. 2.

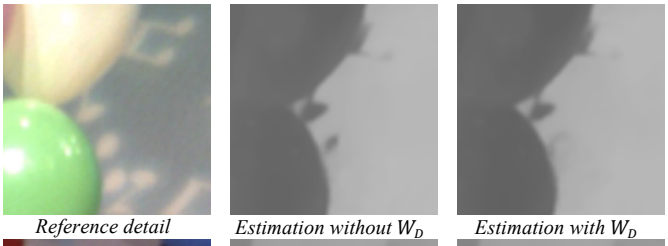


Figure 2. Disparity estimation improvements using (7) and (9) First row: better edge preservation and minimization of foreground disparity values transferred to background elements using w_D Second row: better disparity refinement allowing N_R growth.

In Fig. 3 selected disparity estimations are presented, omitting some processing steps of the algorithm. In the same figure, the consistency check mask M and a final result with all processing blocks activated are illustrated.

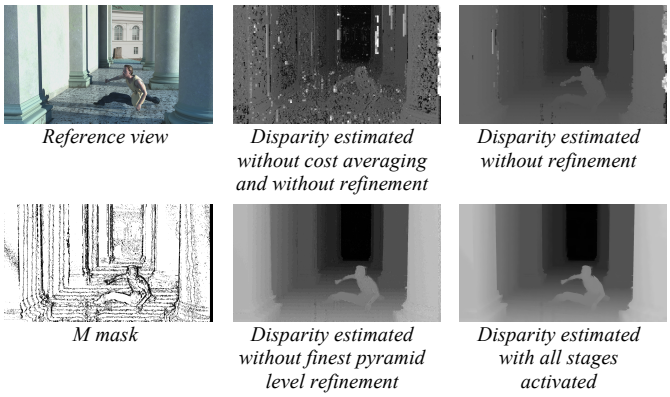


Figure 3. Disparity estimations comparison and consistency check mask.

III. Experimental Results

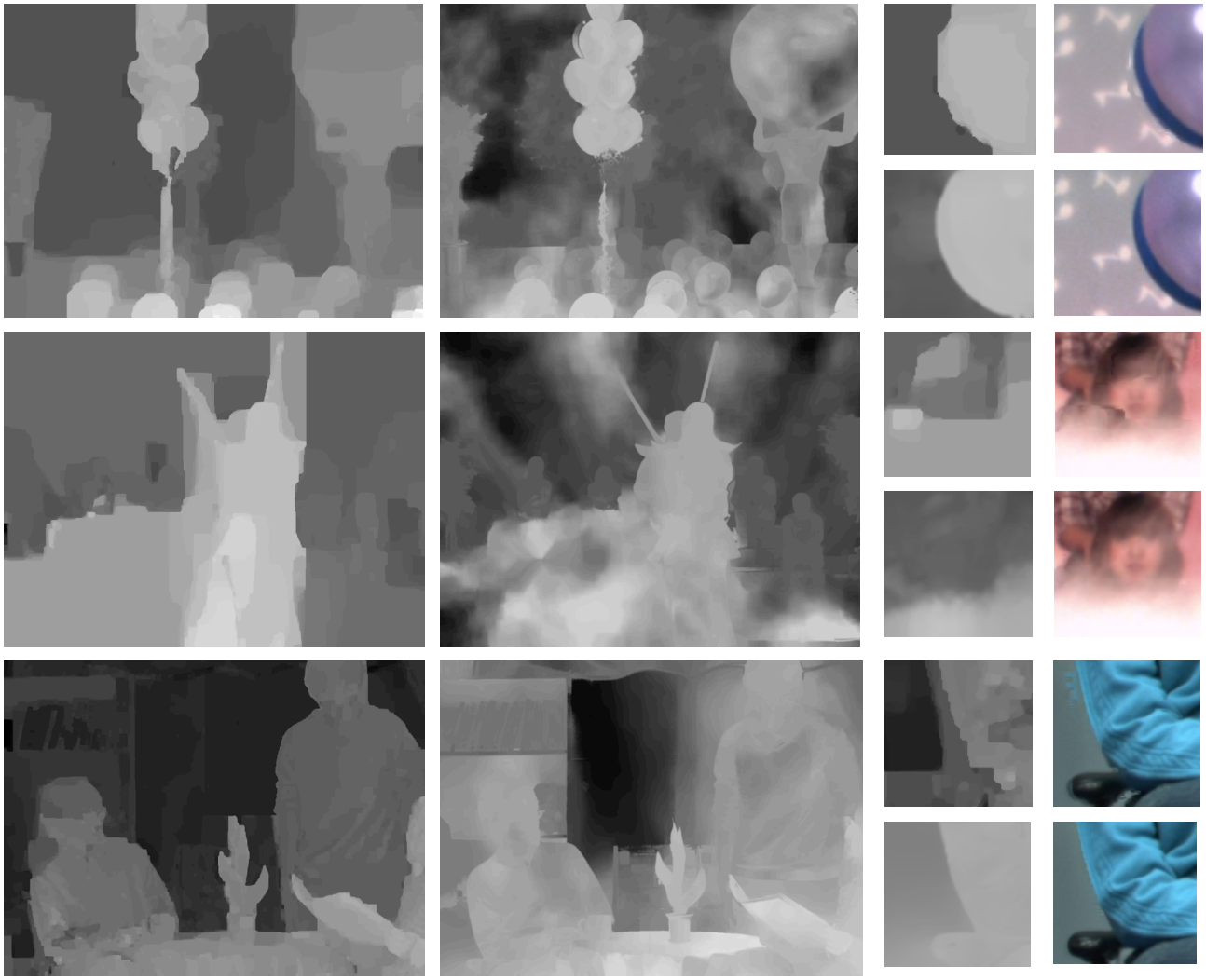
Experimental results are presented as depth maps and intermediate synthetic views. We compare our generated depth maps with the ones provided in [11] which are partially hand-optimized and the synthetic views obtained with our depth against the ones synthesized with the provided depth in [11]. The mentioned synthetic views are generated using the 3D-HEVC renderer described in [12], which requires as input a stereo video and the respective depth map for each view encoded as gray scale videos. Following the description in [12] we convert our dense disparity maps to depth maps. The parameters of our algorithm for all the experiments are set to: R = 2, N_{avg} = 5x5 pixels, N_R = 15x15 pixels, σ_x = 4.2 in averaging and 30 in refinement, σ_c = 20, τ = 0.4, L = 5, β = 0.05 and σ_d = 15.8.

In Fig. 4 we show our depth maps highlighting the accuracy of the estimations near arbitrarily-shaped depth discontinuities and the improvements these estimations produce in the synthetic views. Even considering its minor relevance in the rendering process, given the local approach of the algorithm, it is shown that wrong disparity values can be estimated in texture-less regions.

As a consequence of the hierarchical approach of the algorithm, small search ranges and small cost averaging neighborhoods can be set, making it possible to obtain high quality results in a small amount of time. Furthermore given the parallel nature of the algorithm, the two disparity maps necessary for the 3D-HEVC renderer are generated simultaneously.

IV. Conclusions and Future Work

We have presented a new disparity estimation algorithm capable of generating disparity maps with high accuracy in depth discontinuities. Mapping the estimated disparity to depth maps and using the 3D-HEVC renderer, high quality intermediate views can be obtained. Given the local approach of the algorithm, it is highly parallelizable and can be implemented in GPUs enabling the possibility to obtain



Left column: provided depth maps in [11], center column: generated depth maps with our algorithm, right column: depth and synthesis details: top: detail of provided depth and its correspondent synthesis result. Bottom: generated depth with our algorithm and its synthesis result.

Figure 4. Exemplification of our results.

disparity maps with well-preserved object boundaries for high definition or 4K stereo footage in reasonable time.

V. REFERENCES

[1] Qingxiong Yang, Liang Wang, Ruigang Yang, Henrik, David Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling." CVPR 2006.

[2] Michael Bleyer, Margrit Gelautz, "Graph-based surface reconstruction from stereo pairs using image segmentation", Proc. SPIE 5665, Videometrics VIII, 288, February 28, 2005.

[3] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.", IJCV, 47 (1/2/3):7-42, 2002

[4] S.B. Kang, R. Szeliski, C. Jinxjang, "Handling occlusions in dense multi-view stereo", Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 103-110, 2001.

[5] K. Zhang, J. Lu, G. Lafruit, "Cross-based local stereo matching using orthogonal integral images", IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 7, pp. 1073-1079, Jul. 2009.

[6] K. J. Yoon, S. Kweon, "Adaptive support-weight approach for correspondence search", Proc. IEEE Comput. Vision Patt. Recog., vol. 1., pp. 556-561, Jun. 2003.

[7] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence", CVPR, pp. 1-8, 2008.

[8] Paris, S., Kornprobst, P., Tumblin, J., Durand, F., "Bilateral filtering: theory and applications", Foundations and Trends in Computer Graphics and Vision, vol. 4, pp. 1-73.

[9] S. Mattoccia, M. Viti, F. Ries, "Near real-time fast bilateral stereo on the GPU", ECVW, 2011.

[10] Michael W. Tao, Jiamin Bai, Pushmeet Kohli, and Sylvain Paris. "SimpleFlow: A Non-iterative, Sublinear Optical Flow Algorithm", Eurographics 2012, 31(2), May 2012.

[11] "Call for proposals on 3D video coding Technology, ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036", March 2011, Geneve, Switzerland.

[12] G. Tech, K. Wegner, Y. Chen, S. Yea, "3D-HEVC Test Model 3", MPEG number m28377, January 2013, Geneve, Switzerland.