Abstract
In recent years, new active range sensors have been developed for 3D data acquirement, such as time-of-flight cameras. These sensors enable acquiring of range data at video rate and are suited for dynamic environment. Unfortunately, the resolution of the range data is quite limited and the captured data are typically contaminated by noise.
In this paper, we propose a novel method for depth video enhancement. Using high resolution color video as guidance reference, we iteratively refine the input depth map based on a newly presented linear filter model, in terms of both its spatial resolution and depth precision. The linear filter has a good edge-preserving property and a runtime independent of filter size, which fulfills both accuracy and speed requirements. For temporally consistent estimate on depth video, we extend the method into temporally neighboring frames. Simple optical flow and patch-based similarity measure are used to obtain accurate depth in an efficient manner. Experimental results show that the proposed method greatly improves the quality and boosts the resolution of range data while achieving high computational efficiency. We also show that the temporally consistent constraint addresses a flickering problem and improves the accuracy of depth video.

Key words: Range Data, Depth Video Enhancement, Linear Filter, Temporal Consistency

1. Introduction
Accurate depth at high resolution is required in many applications such as 3DTV, new view rendering, robot vision, etc. A variety of depth measuring methods have been developed, for example, laser range scanners or active illumination with structured lights. However these traditionally accurate depth measuring methods can only be used in static environment. To provide depth maps at video rate for dynamic scenes, passive stereo matching and recent active depth sensors based on time-of-flight (TOF) are two main choices. Unfortunately the quality of depth maps provided by these techniques is often not at a level desired by high-level applications due to the physical limits or real-time constraint. For instance the depth maps obtained by a ToF sensor, ‘Mesa Imaging SR4000’, are of low resolution (176x144) and noisy. This paper hence focuses on a post-processing step that enhances the resolution and quality of given non-ideal depth data.
To resolve the above limits, researchers have presented many methods to enhance the depth data. In filter-based methods, joint bilateral filter is widely used and may work well, because it preserves the discontinuity of up-sampled depth map. Kopf et al.\(^{(1)}\) proposed a general framework for multi-modal image enhancement. In the case of up-sampling spatial resolution, the depth map is up-sampled by joint bilateral filter guided by a color image. Their assumption is the discontinuity of depth maps always coinciding with that of color images. However this assumption is not always valid, such as in high texture region, the joint bilateral up-sampling method admittedly produces copying texture from the color image to real smooth area in the...
depth map. To avoid this problem, Derek Chan et al.\(^{(2)}\) extended the original joint bilateral filter by switching two range filters with a blending function and reduced the influence of color image in those areas where standard bilateral filter is likely to cause erroneous texture copy. Kim et al.\(^{(3)}\) presented a newly designed joint bilateral filter to increase the spatial resolution of depth maps by multiplying two range filters and considering the color and depth information at the same time. Yang et al.\(^{(4)}\) proposed an iterative joint bilateral enhancement method. In contrast to the method\(^{(1)}\) that applies the filter on depth values, the joint bilateral filter is performed on 2D cost slice of each depth candidate and the final depth values are acquired in a WTA (Winner Takes All) framework after a fixed number of iterations. Yang’s method results in better edge-preserving performance, while its computational complexity is \(NumD\) times of 2D joint bilateral up-sampling (JBU) method\(^{(1)}\), where \(NumD\) is the number of depth candidates. The high computational complexity limits JBU use in real-time application. Several methods enable joint bilateral filter to be computed at constant time or even video rate by modifying the model or using GPU implementation\(^{(5),(6)}\). Yang et al. improved the method\(^{(5)}\) and presented an acceleration strategy by sweeping the plane through the intensity level and computing the depth in the intensity order\(^{(7)}\). However these fast methods usually speed up while sacrificing accuracy at the same time.

In the above methods, any other information of neighboring frames in temporal domain are not considered, which may cause the temporal flickering problem. The temporal flickering is an important issue to be addressed. It is usually considered in depth video enhancement. That is not only spatial consistency but also temporal consistency be enforced. One attempt to enforce temporal coherence for depth video enhancement is to smooth the cost volume with a rectangular spatio-temporal support window. For example, Choi et al.\(^{(8)}\) presented a 3D JBU model, in which 2D joint bilateral filter is extended into a 3D volume by accumulating neighboring frames. In contrast to the 3D filter-based method, another attempt is to build cost volume including temporally neighboring frames and then perform the 2D filter on every frame. For example, a new enhancement method for depth video was proposed in the paper\(^{(9)}\) based on a weighted mode filter. And its computational complexity greatly decreases by controlling the width of Gaussian function. These methods cannot cope well with considerably large movement of scene object. To account for faster moving scene object, optical flow field is often computed between consecutive frames and then the depth values is filtered along the computed flow vectors\(^{(9),(10)}\).

In this paper, we proposed a new depth enhancement method based on a recently presented guided image filter\(^{(11)}\). The linear filter has been proven useful and effective in stereo matching\(^{(12)}\) and multi-label filter\(^{(13)}\). The contribution of this paper is to integrate this filter to depth map enhancement task and extend the method to depth video considering temporal consistency. Similar with the 3D bilateral filter method\(^{(6)}\), firstly we build a 3D cost volume based on the current depth values. Then the guided image filter performs on each section of cost volume, and final depth is selected in a WTA framework after a fixed number of iterations. For temporally consistent estimate on depth video, we extend this method to temporally neighboring frames. The simple optical flow and patch-based similarity measure are used to determine the weight of temporally neighboring frame.

The remainder of the paper is organized as follows. In section 2, we describe the depth video enhancement technique using the proposed method. Then the objective and subjective experimental results are given in section 3. Finally we give conclusions in section 4.

2. Depth video enhancement

In general the aim of depth video enhancement is to increase the spatial resolution, suppress the noise and handle the temporal flickering problem. In this paper we present a novel method based on the guided image filter for achieving these goals. Firstly we will simply introduce the linear filter model, then give our proposed depth video enhancement method.
2.1. Linear filter model

We consider a depth enhancement problem, where takes color images and depth map to be processed as inputs. Given a depth map $D$, our goal is to enhance its spatial resolution and quality by using aligned high-resolution color image $I$. The resulting enhanced depth map at the same resolution of $I$ is denoted as $J$. The classical method is based on joint bilateral filter which is a extension of bilateral filter. Formally, for a pixel $p$ in depth map $D$, the enhanced value $J(p)$ is represented by

$$ J(p) = \frac{1}{K_p} \sum_{q \in N_p} W_S(p-q)W_C(I(p)-I(q))D(q) $$

(1)

where $W_C$ and $W_S$ are range filter kernel and spatial filter kernel based on intensity distance and spatial distance of pixel $p$ and $q$ respectively, which both are Gaussian functions with variance $\sigma_C$ and $\sigma_S$ respectively. $N_p$ is a neighborhood around $p$, and $K_p$ is a normalizing factor. Joint bilateral filter is an edge-preserving smooth technique in which the guidance image $I$ and filtered image $D$ are different, which enables edge preserving up-sampling depth maps to the resolution of reference image.

However the computational complexity of its direct implementation is high. Recent some methods for fast implementation have been presented but may sacrifice accuracy. We formulate the depth enhancement problem using the guided image filter proposed recently. The guided filter has a good edge-preserving property and a runtime independent of filter size, which fulfills both accuracy and speed requirements. Thus the state-of-the-art results can be achieved without the need to trade off the accuracy against efficiency. We simply describe the filter model as follows.

In the linear filter model, each pixel $p$ of the output depth map $J$ is supposed to be a linear transform of the guidance image $I$ in a local window $N_k$ centered at a pixel $k$. The local window $N_k$ involves the pixel $p$ and linear transform is expressed by

$$ J(p) = a_k I(p) + b_k $$

(2)

where $a_k$ and $b_k$ are constant parameters of the window $N_k$. The linear relationship ensures the conservation of edges between guidance image $I$ and output image $J$. So we can up-sample the low resolution depth map while preserving the edge of high resolution guidance image. For color image, the above linear transform equation is rewritten as

$$ J(p) = a_k^T I(p) + b_k $$

(3)

where $I(p)$ is a $3 \times 1$ RGB components vector for pixel $p$ and $a_k$ is a $3 \times 1$ coefficient vector. The window parameters $a_k$ and $b_k$ can be determined by minimizing differences between input image $D$ and output image $J$. It has been proven that the window parameters in (3) can be expressed as

$$ a_k = (\Sigma_k + \epsilon U)^{-1} \frac{1}{N} \sum_{q \in N_k} I(q)D(q) - \mu_k \bar{D}_k $$

(4)

$$ b_k = \bar{D}_k - a_k^T \mu_k $$

(5)

where $\Sigma_k$ is the $3 \times 3$ covariance matrix of guidance image $I$ in the window $N_k$, $U$ is a $3 \times 3$ identity matrix, $\mu_k$ and $\bar{D}_k$ are the mean of $I$ and $D$ in $N_k$ respectively, $N$ is the number of pixels in $N_k$, and $\epsilon$ is a parameter to distinguish the smooth and discontinuity areas, which is similar with the range variance of bilateral filter. The detailed analysis of the property of linear filter can be found in the paper.

Then we can apply the linear transform model to all the windows in the entire image. However a pixel $p$ is involved in different windows that contain $p$ which have different window parameters. That is the output for a pixel $p$ is not the same when computed in different
windows. A simple method is to average all the output values of pixel $p$. The final linear model is given by

$$J(p) = \frac{1}{N} \sum_{k \in \Omega} (\bar{a}_k^T I(p) + b_k)$$

$$= \bar{a}_p^T I(p) + \bar{b}_p$$  \hspace{1cm} (6)

where $\bar{a}_p$ and $\bar{b}_p$ are the mean of parameters for $N$ windows involving the pixel $p$. It is noting that the relationship among $D$, $I$ and $J$ given by (6) is also expressed as

$$J(p) = \sum_{q \in \Omega_p} W_{pq}(I)D(q)$$  \hspace{1cm} (7)

where $W_{pq}$ is the kernel weight between pixel $p$ and its neighboring pixel $q$. So the linear model (6) is indeed in the form of image filter and can be compared with the other filters.

2.2. Depth video enhancement algorithm

In this section, we describe our proposed depth enhancement method based on the linear filter. The input images are the low resolution depth map contaminated by noise and the high resolution guidance color image. First, up-sample the depth map to the same size as color image by bilinear interpolation. The interpolated depth map is called initial depth map and not reliable. Then the cost volume is constructed based on current depth values, and the linear filter is performed on each section of cost volume corresponding to every depth candidate to produce new cost volume. The depth value is updated in a WTA framework. After a fixed number of iterations, the above process stops and the final depth map is acquired. The framework of our iterative refinement module for depth map is shown in Fig. 1. It is worth noting that our framework is similar with that of the iterative method\(^{(4)}\). The main difference between the two methods is the cost function and filter model used in depth enhancement framework. In our iterative enhancement method, the cost volume firstly is built according to current depth values. In contrast to the 3D-JBU method\(^{(4)}\), we adopt TAD (Truncated Absolute Difference) as cost function based on distance between current depth $D^k$ and potential depth $d$, where $k$ is the current iteration level and $D^0$ equals the interpolated initial depth map. As we all know, this model can allow initial depth errors. For a pixel $p$, its current cost $C^k(p, d)$ on the depth level $d$ is expressed as

$$C^k(p, d) = \min ((d - D^k(p)), \eta)$$  \hspace{1cm} (8)
where $\eta$ is a predetermined constant. The cost of all pixels composes one section of cost volume corresponding to depth candidate $d$.

In order to apply the proposed method to the depth video, temporal consistency should be considered by using the temporally neighboring frames. Temporally consistent depth estimate from the low quality depth video provides a flicker-free depth video as well as improves the accuracy of an output depth video. An additional computational complexity of temporally consistent estimate should be small compared to the filter process for a single depth map.

In this paper, the temporal neighbors are determined by optical flow method. Recently a number of optical flow algorithms have been presented (14) – (16). However their complexity is still too high to be used in this application. So the simple Horn-Schunck method (17) is applied in this paper. Considering the temporally neighboring estimate may provide erroneous results on depth discontinuities. We hence use the patch based similarity measure $W_t(p, p_n)$ between $p$ in current $t^{th}$ frame and corresponding pixel $p_n$ in temporally neighboring $n^{th}$ frame with the optical flow method together, which is similar to the method (9). The patch similarity measure $W_t(p, p_n)$ is expressed as

$$W_t(p, p_n) = \frac{1}{Z_p} \exp\left(-\frac{\sum_{m} |I_t(p + m) - I_t(p_n + m)|}{\sigma_p}\right)$$

where $Z_p$ is a normalizing factor, $p + m$ and $p_n + m$ are neighbors of $p$ and $p_n$ respectively, whose distances are $m$, and $\sigma_p$ is a gaussian variance. Fig. 2 shows the optical flow and patch-based similarity measure strategy. The final cost volume of the $t^{th}$ frame for the $k^{th}$ iteration is computed through an adaptive summation of cost of temporal neighbors and described by

$$C^k_t(p, d) = \sum_{m \in T_t} W^k_{m}(p, p_n)C^k_m(p_n, d)$$

where subscript $t$ represents the current frame number and superscript $k$ represents the current iteration level, $T_t$ is the neighboring frames of the $t^{th}$ frame.

From the description in section 2.1, we know that the linear filter can smooth the cost space to reduce noise while preserving the edge of color image. Firstly for every frame of depth video, the cost volume is built as described above by using (8),(9) and (10) with temporal consistency. Then the linear filter is performed on 2D cost slice corresponding to each depth candidate using (4),(5) and (6) for every frame. It is worth noting that the linear filter is computed in constant time by means of integer image technique. As a consequence, the computational time of cost filter is independent of the local window size. The new depth is selected according to the minimal filtered cost in depth search range. The above process proceeds until the final depth map is acquired after a fixed number of iterations.

3. Experimental results

To validate the effectiveness and efficiency of the proposed method, we evaluate our method through various experiments. The performance was compared with 2D-JBU method (1) and 3D-JBU method (4). For the three algorithms we use our own Matlab implementation using the Intel Core i3 CPU, 2.3GHZ PC. In this section we perform four different tests. First,
the depth enhancement method is evaluated by applying it to refine depth maps, which are estimated by simple stereo matching algorithm. Then we perform experiments with ground truth depth maps provided by the Middlebury test bed(18). The low resolution depth maps are generated by down-sampling truth depth maps and then up-sampled by our proposed method. Then the down-sampled depth maps in the second test are added additive white Gaussian noise. Last the temporally consistent estimate performance is evaluated. For the first three tests, the proposed method performs on single frame without temporal consistency. That is the cost volume is built using (8) and the performances of up-sampling spatial resolution and suppressing noise are evaluated.

Firstly, the initial depth maps are computed by stereo matching and have the same size as color images, which is different from the following tests. In fact, the stereo matching algorithms compute the disparity values which are inversely proportional to the depth values. For the sake of simplicity, we use only depth for the two types of data in this paper. The proposed method as a refinement step to enhance the initial depth maps. The parameters of our method are: $\epsilon = 0.0001$, window radius $r = 11$, cost truncated threshold $\eta = 0.5 \times \text{NumD}$, where NumD is the number of depth candidates. The same parameters are used for all test images. Fig. 3 shows the initial depth maps and the enhanced ones of Tsukuba provided by Middlebury benchmark(18). From the figures, we can see that the proposed method has significantly improved the initial depth maps, and the algorithm after three iterations has almost converged and the discontinuities are well preserved. By visual comparison, the difference between results after ten iterations and that after three iterations is tiny. Considering speed and effectiveness, the number of iteration is set to 3 in the following experiments.

Secondly, we evaluate the performance of up-sampling low resolution depth maps. The low resolution depth maps are generated by down-sampling the ground truth depth maps provided by Middlebury benchmark(18). The down-sampling ratio is set to 8. The low resolution depth maps are firstly up-sampled to the same size as high resolution color image by bilinear interpolation. The interpolated depth maps are called initial depth maps and then enhanced by our proposed method. The other parameters are the same with those of the depth refinement. The results of our method compared with the 2D-JBU and the 3D-JBU method are given in Fig. 4. The 2D-JBU method performs the joint bilateral filter on initial depth map non-iteratively while the 3D-JBU performs filter on initial cost space built on current depth values iteratively. In order to fairly compare performances of filter-based three methods, we adopt the same initial depth maps and cost function. The number of iteration is 3 for our method and 3D-JBU method. The other parameters of two compared methods are adjusted to acquire the optimal results. From the figures, we can see that the proposed method yields superior results over the two compared methods, especially in discontinuity and occluded areas. The objective evaluation of these methods is shown in Table 1. The accuracy is evaluated by measuring the percent of bad pixels (where the absolute depth error is greater than 1 pixel) for Vis. pixels (that are visible in the image) and Dis. pixels (whose neighboring depth values differ by more than 2).

It is worth noting that the process time of 2D-JBU is the smallest among three methods, but its quality is the worst. In Fig. 5, the execution time of our proposed algorithm and 2D-JBU method on Teddy is represented versus the filter window size. From the figure, we can see that the key of our method is to adopt a linear filter whose execution time is independent of the window size, which makes the algorithm scalable for future applications using higher resolution images.

Thirdly denoising performance of our proposed method is evaluated. The down-sampled depth maps are added additive white gaussian noise with a mean of 0 and standard variation 0.01. Then the proposed method is performed on the noisy initial depth maps and results are shown in Fig. 6. We found that the proposed method may provide accuracy high resolution depth maps even in a noisy environment.

Lastly in order to evaluate the temporally consistent estimate performance of the pro-
posed method, we perform experiments with color video and ground truth depth video of 'Tanks' (400 × 300), provided by\textsuperscript{(19), (20)}. The ground truth depth video is then down-sampled by a factor of 4 and additive white gaussian noise is added with a mean of 0 and standard variation 0.01. The number of neighboring frames $T_f$ is set to 1 and the $t - 1^{st}$ frame is used to compute the final cost volume $C_t(p, d)$ in (10). It is noting that only the previous frame is used here and the cost of the previous frame $C_{t-1}$ is filtered value in (10). The patch size $m$ and the variance $\sigma_P$ are set to 5 × 5 and 40 respectively in (9). The parameters of linear filter are the same with those of the above tests. The enhanced results for 10\textsuperscript{th} frame and 20\textsuperscript{th} are shown in Fig. 7. For objective evaluation, we measure the percent of bad matching pixels with ground truth depth video as given in Fig. 8. The experiment is performed with 40 frames. The temporal consistency is not enforced in the first frame and the results are the same for two compared methods. From the figures, we can see that the results of temporally consistent estimate ('with temporal consistency') are superior to that of depth up-sampling on the single frame ('without temporal consistency'). The patch-based similarity measure can reduce the erroneous estimation of optical flow to obtain temporal smoothing results.

4. Conclusions

In this paper, we have presented a novel approach for low resolution and noisy depth video enhancement. First initial depth maps are built by up-sampling the low resolution depth maps using bilinear interpolation. Then the linear filter is performed on each slice of cost volume based on current depth values. The new depth maps are computed in a WTA framework and used in next iteration. The final depth maps are acquired after a fixed number of iterations. Next the proposed method was extended into the depth video for acquiring improved and temporally consistent results. The temporally neighboring pixels are estimated by the simple optical flow algorithm, and the patch similarity measure is used to reduce the influence of the inaccurate the optical flow estimates. Our iterative framework can effectively enhance the spatial resolution and quality of depth video. Especially the computational complexity based on linear filter model does not depend on the filter size and the proposed method is efficient. In further work, we will design the non-iterative scheme and implement the proposed

Table 1 Objective evaluation for enhanced results

| Methods | Tsukuba | Venus | Teddy | Cones |
|---------|---------|-------|-------|-------|
|         | Vis.    | Dis.  | Vis.  | Dis.  | Vis.  | Dis.  | Vis.  | Dis.  |
| INITIAL |         |       |       |       |       |       |       |       |
| 14.50   | 48.74   | 2.38  | 41.49 | 18.10 | 57.02 | 18.34 | 57.04 |
| 12.20   | 35.70   | 3.30  | 50.45 | 16.13 | 56.85 | 18.52 | 56.91 |
| 2D-JBU  |         |       |       |       |       |       |       |       |
| 12.22   | 36.87   | 2.48  | 42.34 | 14.59 | 49.10 | 17.31 | 49.45 |
| 10.15   | 24.00   | 0.79  | 12.63 | 13.67 | 43.23 | 16.23 | 37.67 |
| PROPOSED |        |       |       |       |       |       |       |       |

Fig. 3 Depth refinement results for Tsukuba:(a) The initial depth map,(b) Depth map after one iteration, (c) Depth map after three iterations, (d) Depth map after ten iterations.
method with GPU for real-time performance.

Acknowledges

Thanks to Support of National Natural Science Foundation of China under Grant Nos. 61202151 and 61202364, the Project of Shandong Province Higher Educational Science and Technology Program under Grant Nos. J12LN34, Program for Scientific Research Innovation Team in Colleges and Universities of Shandong Province.

References

(1) Johannes Kopf, Michael F. Cohen, Dani Lischinski, Matt Uyttendaele, Joint Bilateral Upsampling, ACM Transactions on Graphics, Vol.(26),3,2007.
(2) D. Chan, H. Buisman, C. Theobalt, S. Thrun, A Noise-Aware Filter for Real-Time Depth Upsampling, Proc. of ECCV Workshop on Multicamera and Multi-modal Sensor Fusion Algorithms and Applications, 2008.
(3) Sung-Yeo Kim, Ji-Ho Cho, Andreas Koschan, Mongi A. Abidi, Spatial and Temporal Enhancement of Depth Images Captured by a Time-of-flight Depth Sensor, Proc. of IEEE International Conference on Pattern Recognition,pp.2358-2361, 2010.
(4) Yang Q., Yang R., Davis J., Nister D., Spatial-depth Super Resolution for Range Images, Proc. of IEEE Computer Vision and Pattern Recognition, 2007.
(5) Chen J., Paris, S., Durand F., Real-time Edge-aware Image Processing with the Bilateral Grid, ACM Transactions on Graphics, Vol.(26),3,2007.
(6) Xueqin Xiang, Guangxia Li, Jing Tong, Mingmin Zhang, Zhigeng Pan, Real-Time Spatial and Depth Upsampling for Range Data, Transactions on computational science, Vol.(6670), pp.78-97, 2011.
Fig. 5 Execution time of two compared algorithms for Teddy.

Fig. 6 Depth up-sampling results in a noisy environment for "Tsukuba, Venus, Teddy and Cones": (a) The input reference color images, (b) The initial depth maps, (c) The noisy initial depth maps, (d) Up-sampled results by our proposed method.

(7) Yang Q., Tan H.-H., Ahuja N., Real-time O(1) Bilateral Filtering, Proc. of IEEE Computer Vision and Pattern Recognition, pp.383-387, 2009.

(8) Jinwook Choi, Dongbo Min, Donghyun Kim, Kwanghoon Sohn, 3D JBU Based Depth Video Filtering for Temporal Fluctuation Reduction, Proc. of IEEE International Conference on Image Processing, pp.2777-2780, 2010.

(9) Dongbo Min, Jiangbo Lu, Minh N. Do, Depth Video Enhancement Based on Weighted Mode Filtering, IEEE Transactions on Image Processing, Vol.(21), 3, 2011.

(10) Bleyer M., Gelautz M., Temporally Consistent Disparity Maps from Uncalibrated Stereo Videos, Proc. of the 6th International Symposium on Image and Signal Processing and Analysis, pp.383-387, 2009.
Fig. 7 Temporal consistency in depth video for “Tanks” where the top line is for the 10th frame and the bottom line is for the 20th frame: (a) The input reference color images, (b) The initial depth maps, (c) The enhanced results by the method without temporal consistency, (d) The results by our proposed method with temporal consistency.

Fig. 8 Error matching percents of all pixels on up-sampled depth video.

(11) K. He, J. Sun, X. Tang, Guided Image Filtering, Proc. of IEEE European Conference on Computer Vision, pp.1-14, 2010.
(12) Leonardo De-Maeztu, Stefano Mattoccia, Arantxa Villanueva, Rafael Cabeza, Linear Stereo Matching, Proc. of IEEE International Conference Computer Vision, 2011.
(13) Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, Margrit Gelautz, Fast Cost-Volume Filtering for Visual Correspondence and Beyond, Proc. of IEEE Computer Vision and Pattern Recognition, 2011.
(14) D. Sun, E. Sudderth, M. Black, Layered Segmentation and Optical Flow Estimation over Time, Proc. of IEEE Computer Vision and Pattern Recognition, 2012.
(15) L. Chen, J. Wang, Y. Wu, Decomposing and Regularizing Sparse/non-sparse Components for Motion Field Estimation, Proc. of IEEE Computer Vision and Pattern Recognition, 2012.
(16) http://vision.middlebury.edu/flow
(17) B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence, pp.185-203, 1981.
(18) http://vision.middlebury.edu/stereo
(19) C. Richardt, D. Orr, I. Davies, A. Criminisi, N. A. Dodgson, Realtime Spatiotemporal Stereo Matching Using the Dual-Cross-Bilateral Grid, Proc. of IEEE European Conference on Computer Vision, 2010.
(20) http://www.cl.cam.ac.uk/research/rainbow/projects/dcbgrid/datasets