Memory Efficient Stereoscopy from Light Fields

Changil Kim[1,2] Ulrich Müller[1,2] Henning Zimmer[1] Yael Pritch[1]
Alexander Sorkine-Hornung[1] Markus Gross[1,2]

[1]Disney Research Zurich [2]ETH Zurich

Abstract

We address the problem of stereoscopic content generation from light fields using multi-perspective imaging. Our proposed method takes as input a light field and a target disparity map, and synthesizes a stereoscopic image pair by selecting light rays that fulfill the given target disparity constraints. We formulate this as a variational convex optimization problem. Compared to previous work, our method makes use of multi-view input to composite the new view with occlusions and disocclusions properly handled, does not require any correspondence information such as scene depth, is free from undesirable artifacts such as grid bias or image distortion, and is more efficiently solvable. In particular, our method is about ten times more memory efficient than the previous art, and is capable of processing higher resolution input. This is essential to make the proposed method practically applicable to realistic scenarios where HD content is standard. We demonstrate the effectiveness of our method experimentally.

1. Introduction

Stereoscopic 3D content creation and manipulation have received much attention in the computer vision and graphics research communities as well as the entertainment industry, which motivated the development of various novel acquisition, processing, and display technologies. Still, stereopsis is a complex function of parallax, inter-axial distance, screen size, viewing distance and more [26], rendering it difficult to create content that provides a comfortable viewing experience while suitable for different viewing preferences and conditions. Changing the inter-axial or the convergence modifies the depth perception *globally* and hence provides only limited control over the disparity distribution, often resulting in over-flattening with most local details vanished. These challenges inspired a considerable body of work on *local* stereoscopic content editing [2, 5–8, 13, 16, 22, 27]; see [19] for a more complete review.

A number of methods among them take a monoscopic view and create the second view given desired depth criteria [2, 5, 27], usually in the form of a target disparity map or other forms of annotation, while other methods warp a given second view to meet the criteria [4, 6, 16]. Most of these techniques employ image inpainting or image warping to realize the intended binocular parallax and fill in the previously unseen, but now disoccluded monocular regions. However, this essentially *hallucinates* potentially incorrect image content. Furthermore, in most cases it is not clear how to extend these methods to utilize more than two views to improve results even in cases where more image information is available. For these reasons, these approaches are subject to limitations with respect to the fidelity and the amount of disparity change that are achievable, and often exhibit noticeable image deformation as the target disparity diverges from the original disparity.

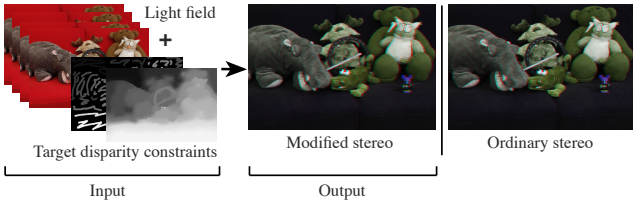


Figure 1. Our methods takes a light field and target disparity constraints the user wants to achieve, and synthesizes the stereo pair which best fulfills the desired constraints. It allows the user to modify the depth perception in a more expressive and robust way. This example shows a use case based on scribbles, where the relative depths of the hippo and the bear are reversed; compare to the ordinary stereo pair formed of two views from the input light field.

A different class of methods is based on the observation that depending on the desired disparity constraints, the resulting second image is effectively a *multi-perspective* image, i.e., it purposely deviates from a standard perspective image in order to optimize the stereo impression for given constraints [13, 22]. The methods of this sort use light fields that already retain multiple perspectives, and thus have more degrees of freedom to choose the most appropriate image content to composite the resulting stereo image pairs. While they provide finer local control over the resulting stereo and do not suffer from image distortions, they are more memory intensive than those using fewer views. Further, most dis-

parity editing methods require correspondence information between input images (e.g., scene depth), and the computational burden to find correspondences becomes significant with an increasing number of input views.

In this paper, we propose a novel method to generate stereoscopic content from light fields, which avoids the problems mentioned above. Given any view of the input light field, our method synthesizes a second view that adheres to desired target disparity constraints (see Figure 1). The method allows us to describe the target disparity on a per-pixel basis, and it chooses the correct light rays according to the target disparity, instead of deforming the input images. We formulate this as a variational optimization problem and solve it via primal-dual iterations.

Our method has the following beneficial properties compared to the previous work: 1) it takes multiple input views to handle occlusions and disocclusions properly and thus avoid image distortion, 2) although it supports per-pixel disparity control, it does not require dense correspondence information such as per-view scene depth. 3) it runs under significantly less memory overhead than the previous state of the art, and is thus applicable to high resolution input, and 4) additionally it inherits the advantage of the iterative continuous solver, such as interactive update and no grid bias.

2. Related Work

We briefly review the existing techniques about stereoscopic content editing roughly in the order of increasing expressiveness. We refer the reader to Masia et al.'s survey [19] for an overview of a broader range of related techniques.

Our method is also inspired by multi-perspective imaging and light field rendering [17]. Most importantly, Seitz [25] analyzes the space of all possible stereo image types and shows that the stereoscopic depth perception using multi-perspective images is feasible.

Stereoscopic Camera Control. The most basic means of disparity modification is to change the inter-axial distance, the distance between the two cameras' optical centers, and the convergence, the amount of rotation against each other around their vertical axes [16]. In recent work, the control of these rig parameters is almost fully automated according to the content of the scene about to be captured or rendered. Heinzle et al.'s computational stereo camera [12] analyzes the scene it is capturing in real-time and adjusts those parameters as well as others, so that the captured scene remains in the comfort zone. Oskam et al. [21] implements a similar idea in the context of real-time rendering such that the virtual scene is rendered in a fail-safe manner. Koppal et al. [15] explored the space of stereo parameters extensively in their production pipeline. Albeit the strong modification capability and intuitiveness of these parameters, their expressive power is notably limited in that their change introduces a *global* effect on the perception of the entire scene geometry, not the *local* control over disparity that we are interested in, and that, more significantly, they cannot be modified once captured.

Stereoscopic Rendering. Besides capturing the content stereoscopically, many techniques that create stereoscopic rendering from the original 2D content have been developed. For CG content, the scene depth is usually given, and thus used to synthesize two or more views for the target display. Bowles et al. [2] proposed a fast image warping technique using fixed point iteration that is ideally targeted for real-time applications such as video games. Being part of the rendering pipeline, it has access to almost the complete information about the scene including the geometry, and the information additionally required can be rendered on demand. Along a similar line is Didyk et al.'s method [5], where the method takes a single image and a depth buffer, and generates two views for the left and right eye using image-space adaptive grid warping at an interactive rate. Masia et al. [18] extend it to generate multiple output views to feed autostereoscopic displays. They also present a perceptually based disparity remapping that can compensate for the limited disparity bandwidth of such displays. Both methods uses the same rendering technique, which handles disocclusions by stretching grid quads and may lead to visual artifacts. While usually available in the animation pipeline for those method, dense depth is not generally given in real-world content. Wang et al. [27] propose an interactive user interface for the creation and manipulation of stereo content, based on sparse user scribbles to annotate the scene depth, which are propagated to fill the entire image space at an interactive rate. However, in their method the resulting images are essentially warped versions of the original image, and thus often include noticeable distortions around the occlusion boundaries in particular.

Local Disparity Editing. For finer control of stereo depth perception of the existing 3D content, the usual strategy is to locally manipulate the disparities of matching image features between the two views. Lang et al.'s method [16] computes sparse correspondences between given two images and warps the images using a variational framework such that the correspondences will have modified parallax in the deformed image pair. To describe the desired artistic manipulation, they formally define a collection of disparity remapping operations, including nonlinear remapping, which enable sophisticated control over the design of disparity modification. Chang et al. [4] allow the user to interact the similar editing process of their method. They use the image warping technique based on 2D mesh deformation to render the output stereo. Focusing on perceptual issues, Didyk et al. [6–8] propose remapping operators and their

implementations that minimize the discomfort perceived by the human visual system. The modified disparity is rendered back to stereo views using the technique based on image decomposition. As for the single view methods, all of these methods use image based techniques to realize modified disparity constraints, which makes them vulnerable to the undesirable extreme deformation or stretching of image content in the course of disparity modification. This limitation could be alleviated by using more input views, but it is not clear how to incorporate such additional information into these methods.

Multi-perspective Approaches. While all above methods take only one of two views, Peleg et al.'s method [22] takes a video cube that captures a 360° panorama, and constructs two views that form a stereoscopic panorama with the disparity locally manipulated. Since their method extracts and composites image columns from the video cube, however, only per-column disparity control is possible. To cope with this limitation, Kim et al. [13] proposed a method that realizes per-pixel disparity control using a light field, but their dense graph-cut formulation poses significant overhead on the computation, in particular memory consumption, preventing it from wider adoption. While Richardt et al.'s Megastereo [24] provides a remedy for the panoramic stereo so that the technique can be applied to HD content, no equivalent for the light field stereo exists so far.

3. Formulation

Our method takes as input a light field and user-defined target disparity criteria. For a given reference view within the light field, our method computes a new view such that the disparity between the views matches best the prescribed target disparity. For our problem, a light field with horizontal angular variation suffices as only the horizontal parallax matters in stereopsis. Such light fields can be easily captured with a 1D camera array or a linear camera gantry. The target disparity is in the form of a 2D map defined at the reference view. We demonstrate a few ways to obtain it in Section 5.

Let $\Omega \subset \mathbb{R}^2$ be the spatial domain of the (continuous) light field, and $\Gamma = [s_{\min}, s_{\max}] \subset \mathbb{R}$ be its bounded 1D angular domain. We then define a light field $L\colon [\Omega \times \Gamma] \to \mathbb{R}^3$, which maps a ray defined by a spatio-angular coordinate $(\mathbf{x}, s)$, where $\mathbf{x} = (u, v) \in \Omega$ and $s \in \Gamma$, to a sampled radiance represented in RGB color space. Further let $\hat{s} \in \Gamma$ denote the position of the reference image $I_{\hat{s}}(\mathbf{x}) = L(\mathbf{x}, \hat{s})$ for which the target disparity map $G\colon \Omega \to \mathbb{R}$ is specified.

In the first step we shift the reference image $I_{\hat{s}}$ by the target disparity G to obtain a *target image*

$$I^*_{\hat{s}}(u + G(u, v), v) = I_{\hat{s}}(u, v) \; . \tag{1}$$

This target image represents what the sought second view should look like. However, as the shifting is not injective nor surjective, there are ambiguities. We deal with the non-injectiveness that would map two pixels to the same location by selecting the pixel with the highest disparity, i.e. the one closest to the camera. To deal with the non-surjectiveness that leaves certain pixels without a disparity value, we mark these undefined regions in a binary mask $M\colon \Omega \to \{0, 1\}$ that is 0 in the undefined regions and 1 elsewhere. The undefined region is the disoccluded, monocular region which, in principle, should not be crucial to the depth perception, but may cause discomfort when conveying conflicting depth cues [16]. Thus, many techniques fill this region by stretching neighboring image regions. However, this often introduces unwanted visible distortions of the image content.

Our Approach. We propose a different approach where we use pixels from the input light field to fill in information in the disoccluded regions. The unknown second view will hence be defined by a labeling function $l\colon \Omega \to \Gamma$ that determines for each pixel in the second view from which input view it should be taken from. To find a smooth solution with a least noticeable transitions (seams) we formulate the problem of finding l as a continuous optimization problem consisting of a data matching and a smoothness term

$$E(l) = \int_\Omega E_{\text{data}}(l) \; + \; k \, E_{\text{smooth}}(l) \; \mathrm{d}\mathbf{x} \; , \tag{2}$$

where $k > 0$ balances the two terms.

The data term E_{data} enforces the resulting second image to be as close as possible to the target image in the subset of Ω where the target image is defined, i.e. where $M(\mathbf{x}) = 1$. Thus the data term is defined as

$$E_{\text{data}}(l) = M(\mathbf{x}) \, \| L(\mathbf{x}, l(\mathbf{x})) - I^*_{\hat{s}}(\mathbf{x}) \|_1 \; . \tag{3}$$

The smoothness term E_{smooth} penalizes the amount of view transitions in the labeling. Importantly, it also guides the transitions to happen in less noticeable regions to allow for a seamless stitching of contributions from different images. For the disoccluded regions where the data term is disabled, the smoothness term allows to fill in information in a smooth manner, resulting in a least distorted completion of these missing regions. To achieve these goals, we define the smoothness term as the anisotropic total variation regularizer [10, 20]

$$E_{\text{smooth}}(l) = \sqrt{\nabla l(\mathbf{x})^\top \, S(\mathbf{x}, l(\mathbf{x})) \, \nabla l(\mathbf{x})} \; . \tag{4}$$

The anisotropy is driven by the local variation in the light field, and measured using the structure tensor [9]

$$S(\mathbf{x}, s) = K_\sigma * (\nabla_{\mathbf{x}} L(\mathbf{x}, s) \, \nabla_{\mathbf{x}} L(\mathbf{x}, s)^\top) \; , \tag{5}$$

where K_σ denotes a Gaussian kernel of variance σ^2, $*$ is the convolution operator, and $\nabla_{\mathbf{x}} L = (\partial_u L, \partial_v L)^\top$ is the

spatial gradient of the light field L. The two orthonormal eigenvectors of S point along and across dominant spatial edges in the light field. Hence our smoothness term aligns view transitions with discontinuities in the light field which minimizes visible seam artifacts due to view transitions.

Plugging above definitions into (2) we end up with the following variational problem to find the sought labeling l:

$$\min_{l} \int_{\Omega} M(\mathbf{x}) \, \|L(\mathbf{x}, l(\mathbf{x})) - I^*_{\hat{s}}(\mathbf{x})\|_1 + k \sqrt{\nabla l(\mathbf{x})^\top \, S(\mathbf{x}, l(\mathbf{x})) \, \nabla l(\mathbf{x})} \, \mathrm{d}\mathbf{x} \, . \quad (6)$$

Convex Formulation. While the regularizer of the functional (6) is convex, the data term is not. We reformulate (6) as a convex functional using function lifting. We only outline the fundamental steps of the procedure here and refer to [23] for more details.

Let us define a binary function $\phi \colon [\Omega \times \Gamma] \to \{0, 1\}$ with

$$\phi(\mathbf{x}, s) = \begin{cases} 1 & \text{if } l(\mathbf{x}) > s \\ 0 & \text{otherwise} \end{cases} \, , \quad (7)$$

which is the indicator for the s-superlevel sets of l. The feasible set of functions ϕ is $\{\phi \colon [\Omega \times \Gamma] \to \{0, 1\} \mid \phi(\mathbf{x}, s_{\min}) = 1, \, \phi(\mathbf{x}, s_{\max}) = 0\}$. Rewriting (6) with ϕ will now yield a convex data term, yet the feasible set of ϕ is non-convex, and hence the minimization over it. To cope with this, ϕ is further relaxed so that it may take continuous values in the interval $[0, 1]$, leading to the convex feasible set

$$D = \{\phi \colon [\Omega \times \Gamma] \to [0, 1] \mid \phi(\mathbf{x}, s_{\min}) = 1, \, \phi(\mathbf{x}, s_{\max}) = 0\} \, . \quad (8)$$

When ϕ is projected back to its original domain after the optimization, it is thresholded by some value within the interval $[0, 1]$. The optimality is still guaranteed regardless the selection of threshold [23]. The labeling function l is recovered from ϕ by integrating over Γ [3]:

$$l(\mathbf{x}) = s_{\min} + \int_{\Gamma} \phi(\mathbf{x}, s) \, \mathrm{d}s \, . \quad (9)$$

Rewriting (6) using the partial derivative of the indicator function ϕ, we obtain the following convex problem:

$$\min_{\phi \in D} \int_{\Omega} \int_{\Gamma} M(\mathbf{x}) \, \|L(\mathbf{x}, l(\mathbf{x})) - I^*_{\hat{s}}(\mathbf{x})\|_1 \, |\partial_s \phi(\mathbf{x}, s)| + k \sqrt{\nabla_{\mathbf{x}} \phi(\mathbf{x}, s)^\top S(\mathbf{x}, s) \nabla_{\mathbf{x}} \phi(\mathbf{x}, s)} \, \mathrm{d}s \, \mathrm{d}\mathbf{x} \, . \quad (10)$$

4. Optimization

A straightforward way to minimize the convex energy functional (10) would be to solve its associated Euler-Lagrange differential equation [23]. This appraoch is, however, complicated by the singularity of the used norms at zero. As an alternative we rewrite the norms in terms of their Wulff shape as the combined two norms constitute a convex, positively 1-homogeneous function [28].

Saddle-Point Formulation. A Wulff shape is defined as

$$W_\phi = \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{y}, \mathbf{z} \rangle \leq \phi(\mathbf{z}) \; \forall \mathbf{z} \in \mathbb{R}^n\} \, , \quad (11)$$

for a convex function $\phi \colon \mathbb{R}^n \to \mathbb{R}$ that is positively 1-homogeneous, i.e. $\phi(\lambda \mathbf{z}) = \lambda \, \phi(\mathbf{z})$, $\forall \lambda > 0$. It is a closed and bounded convex set containing zero, and is used to rewrite ϕ as

$$\phi(\mathbf{z}) = \max_{\mathbf{y} \in W_\phi} \langle \mathbf{z}, \mathbf{y} \rangle \, , \quad (12)$$

where the norms can be represented in a differentiable form. The minimization problem (10) can then be rewritten as

$$\min_{\phi \in D} \max_{\mathbf{p} \in C} E(\phi, \mathbf{p}) \, , \quad (13)$$

with the energy functional

$$E(\phi, \mathbf{p}) = \int_{\Omega} \int_{\Gamma} \langle \nabla_{\mathbf{x},s} \, \phi(\mathbf{x}, s), \, \mathbf{p}(\mathbf{x}, s) \rangle \, \mathrm{d}s \, \mathrm{d}\mathbf{x} \, , \quad (14)$$

where $\nabla_{\mathbf{x},s}$ is now the gradient over all three dimensions of ϕ and $\mathbf{p} = (p_{\mathbf{x}}, p_s)^\top$ is the dual variable. The feasible set of the dual $\mathbf{p}$ then becomes the following Wulff shape:

$$C = \{\mathbf{p} \colon [\Omega \times \Gamma] \to \mathbb{R}^3 \mid \sqrt{p_{\mathbf{x}}^\top \, S(\mathbf{x}, s) \, p_{\mathbf{x}}} \leq k, \, |p_s| \leq \rho(\mathbf{x}, s)\} \, , \quad (15)$$

where $\rho(\mathbf{x}, s)$ is the data term value at $(\mathbf{x}, s)$. This can be seen as a partial dualization in convex analysis, where ϕ is referred to as the primal variable and $\mathbf{p}$ the dual. Because we will maximize in the dual $\mathbf{p}$ and minimize in our original primal ϕ, the problem (13) is called the saddle-point formulation.

Primal-Dual Iterations. To solve (13), we alternate between taking gradient steps in the primal and dual [11]. To minimize the primal, we define the gradient as

$$\frac{\phi^n - \phi^{n+1}}{\sigma_p} = \nabla_\phi E(\phi, \mathbf{p}) \, , \quad (16)$$

and to maximize the dual, we define the gradient as

$$\frac{\mathbf{p}^{n+1} - \mathbf{p}^n}{\sigma_p} = \nabla_{\mathbf{p}} E(\phi, \mathbf{p}) \, . \quad (17)$$

By calculating the derivative of (13) with respect to the primal and the dual we derive the update steps

$$\textbf{Primal:} \quad \phi^{n+1} = \mathcal{P}_D(\phi^n + \sigma_p \operatorname{div} \mathbf{p}^n) \, , \quad (18)$$

$$\textbf{Dual:} \quad \mathbf{p}^{n+1} = \mathcal{P}_C(\mathbf{p}^n + \sigma_p \nabla \phi^{n+1}) \, , \quad (19)$$

where $\mathcal{P}_D$ projects ϕ back into its domain D by truncating it to $[0, 1]$ and setting $\phi(\mathbf{x}, s_{\min}) = 1$ and $\phi(\mathbf{x}, s_{\max}) = 0$. $\mathcal{P}_C$ is the Euclidean projector of the set C given by [23]

$$\mathcal{P}_C(\mathbf{p}^{n+1}) = \arg\min_{\mathbf{y} \in C} \|\mathbf{p}^{n+1} - \mathbf{y}\| \, . \quad (20)$$

(a) User scribbles (b) Propagated disparity (c) Reference image (d) Computed new view (e) Resulting stereo

Figure 2. *Disparity modification using user scribbles.* This task demonstrates a possible use case, where sparse brush strokes are drawn by the user and then propagated to form a dense target disparity map. The resulting stereo is generated according to the propagated disparity map. (a) shows the input scribbles and (b) the resulting target disparity map. (c) and (d), respectively, show the reference view and the computed new view. (e) shows the resulting anaglyph stereo image. Note that the scribbles are not necessarily physically meaningful and are rather intended to test the flexibility and robustness of our method. See the supplementary material for more results.

To compute the updates numerically, we discretize Ω and Γ so they represent pixel coordinates and the image index in the light field, respectively. The gradients are approximated using forward differences, but we use backward differences for the divergence to ensure convergence.

5. Experimental Results

In this section we evaluate our method both qualitatively and quantitatively. We begin with the demonstration of two different use cases: disparity modification using sparse user scribbles, and nonlinear disparity remapping. We then assess our method quantitatively, and finally analyze its performance. The datasets used are taken from two public light field repositories accompanied with [14, 29], and resized to a small set of representative resolutions. For all experiments, we used a fixed set of parameters, $\sigma_p = 1/\sqrt{3}$, $k = 10$, and $\sigma = 2$, and the primal-dual steps were iterated 10,000 times. Due to the limited space, we present only a selected subset of our results in the paper, which are also compressed for smaller file size. For the complete set of higher resolution results, refer to the accompanied supplementary material. All anaglyph images shown can be viewed in 3D using red-cyan anaglyph glasses.

Qualitative Evaluation. Our first use case based on *user scribbles* demonstrates a pipeline for the stereo editing and the 2D-to-3D conversion (see Figure 2). A sparse disparity annotation is provided by the user by drawing several brush strokes on top of the reference image, where the grayscale intensity of strokes encodes the amount of disparity. This sparse input is then propagated to form the dense target disparity map using a standard technique, e.g., StereoBrush [27]. Note that these scribbles do not necessarily need to be physically correct: our method finds the labeling that is closest to the specified disparity while producing the least noticeable seams, which leads to convincing stereo images. For instance, in the scribbles for the Couch dataset, the hippo is pushed back further than all other stuffed animals, while it is the closest in real depth; compare against the ordinary, perspective stereo shown in Figure 1.

Figure 4 shows the second use case, where the actual scene depth is *nonlinearly remapped* to convey a different depth perception. We used the scene depth for the Bikes, Couch, and Mansion datasets that is available in the same light field repository. Since depth is not available for the Elephant dataset, we do not show the corresponding results in the paper. For the Bikes dataset shown in the first row, the depth of the ground is compressed to give more disparity

(a) Ordinary stereo (b) User scribbles (c) Nonlinear remapping

Figure 3. *Side-by-side comparisons between ordinary stereo and our results.* (a) the ordinary stereo consisting of two perspective images chosen from the input light field. (b–c) our results using the user scribbles and the nonlinear disparity remapping, respectively.

(a) Original disparity (b) Remapped disparity (c) Reference image (d) Computed new view (e) Resulting stereo

Figure 4. *Nonlinear disparity remapping.* The actual scene depth of the reference view is nonlinearly remapped to create the target disparity map. (a) shows the original disparity map, whereas (b) depicts the remapped disparity map. For the Bikes dataset, the excessive disparity on the ground was compressed for a more comfortable stereoscopic viewing experience. For the Couch dataset, the gradient of the disparity is modified such that large disparity gradients are removed, to better distribute disparity and to obtain more local details. (c–e) show the reference image, the computed new view, and the resulting anaglyph stereo rendering, respectively.

budget to the bikes at farther distance. For the Couch and Mansion datasets, the disparity gradient is obtained from the disparity map, and high gradient magnitude are truncated to remove empty space and give more local detail. The target disparity map is then reconstructed from the modified gradients using a Poisson solver [1]. Note that such remappings are not realizable by changing the inter-axial distance and/or the convergence.

Figure 1 and 3 present the side-by-side comparisons of the "ordinary" stereo consisting of two perspective images, and our results using user scribbles and nonlinear remapping.

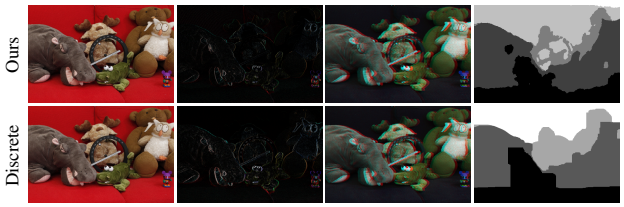
Quantitative Evaluation. To assess our results quantitatively we conducted two experiments where the desired result is known a priori. First, we use a single *constant disparity* for all pixels as the target disparity. Thus our method should result in the same image that is only translated by the amount of the disparity value, by best combining the pixels from the different views. Second, we use a *linearly scaled disparity map* of the reference view as the target disparity. Since this linear scaling does not involve any local disparity modification, our method should choose the same, single input image entirely.

Figure 5 shows the results of constant disparity, where the target disparity is set to 20 pixels for all datasets. The third column shows the absolute difference between the image computed by our method and the reference image translated by the amount of disparity. The resulting anaglyph stereo images which are shown in the next column should look flat, but floating on the screen. The last column shows the resulting labeling, where each step in the grayscale denotes an image index. The labeling resembles the scene depth, and in fact, the stereoscopic rendering problem we are addressing and the dense depth estimation problem are tightly related. See the supplementary material for the detail.

The results of a linearly scaled disparity are shown in Figure 8. We scaled the given depth map at the reference view by a factor of 10, hence each resulting view should equal to its tenth next view. As for the constant case, we show the reference image, the computed new view, and the difference between the computed image and the corresponding input image. The labeling shown in the last column should look close to flat for this experiment.

We measured the root-mean-squared errors (RMSE) of the computed views from the ground truth for all data sets at two different resolutions: 1280×853 (1k) and 1920 ×1280 (2k). For both tasks of the Bikes and Couch datasets, the RMSE was all below 0.04. The error was higher for the Mansion dataset for both tasks, primarily due to the complex and thin structure of the tree and fence, which was about 0.07. Since the Elephant dataset is only available at 1k resolution and no depth was available, we performed the constant disparity task at 1k resolution, which showed the RMSE of 0.05. See the supplementary material for the complete results.

Comparisons and Performance. We show the advantage of our method against the current state of the art [13], which solves a similar problem using a discrete graph-cut formula-



(a) Computed view (b) Error (c) Stereo (d) Labeling

Figure 6. *Comparison to the discrete formulation.* We compare the constant disparity task (see Figure 5) against the start of the art method which uses a discrete graph-cut formulation [13]. The labeling of the discrete formulation clearly shows grid bias, i.e., the transitions are mostly axis-aligned or diagonal (bottom (d)). This results in a higher error (bottom (b)).

(a) Reference image (b) Computed new view (c) Error (d) Anaglyph stereo (e) Labeling

Figure 5. *Constant disparity.* (a) and (b) show the reference image and the computed new view given the fixed value of 20 pixels as the target disparity. (c) shows the error of the computed image against the ground truth, for which we use the reference image translated by 20 pixels (the darker the pixel, the smaller error). (d) shows the anaglyph stereo image, while (e) shows the resulting labeling, where each step in the grayscale denotes an image index. The resulting stereo should ideally look flat, but floating on the screen.

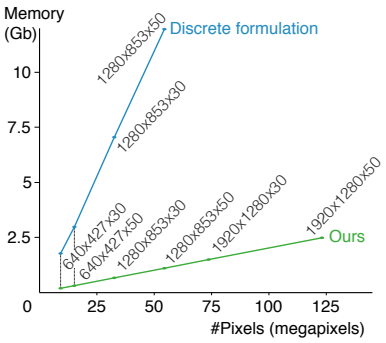


Figure 7. *Memory usage.* This graph shows the amount of memory our method requires for the input with different resolutions. Compared to the state of the art [13], our method is more memory efficient. We were not able to run the discrete formulation for 2k resolutions.

tion. A characteristic problem of the discrete formulations is that the optimization depends on the discretization, which is known as the grid bias. Figure 6 shows a side-by-side comparison of the constant disparity task, where we applied the discrete formulation to reproduce the same task. As seen in the labeling image, the discrete solver yields the labeling that is mostly aligned along the two image axes, and also exhibits a higher error in the final rendering.

We implemented the primal-dual iterations on GPU using NVidia CUDA. The maximal GPU memory that the implementation requires at a time was measured for several different resolutions, both spatially and angularly. We show the memory footprint in Figure 7, also with the comparison to the state of the art method [13]. Our method uses less than 10% of the memory compared to theirs. The running time varies depending on both the type of tasks and the light field resolution. Measured on an Intel i7 processor with 16 GB of RAM and an NVidia GTX 560 graphics card, the running time of the tasks for 50 1k images varied between 10 and 12 minutes and for 30 2k images between 13 and 15 minutes. We refer the reader to the supplementary material for the complete analysis of our method's performance for all results we report in the paper.

6. Conclusion

We presented a method to create stereoscopic 3D content from light fields. Having a light field as input, our method can handle occlusions and disocclusions more properly without deforming the input image. At the same time, the method is more memory efficient, enabling the high resolution input to be processed. Unlike similar methods, ours does not require dense depth information as input, which often turns out challenging to be computed. We formulated this problem as a variational optimization problem, which allows us to avoid the undesired artifact of the discrete formulation such as grid bias or excessive memory consumption.

Fast feedback is essential to an interactive editing. The current running time of our method indicates that follow-up work would be expected to speed up the method. A promising property of our method in this regard is that our solver runs iteratively, and each intermediate solution can be visualized simultaneously to provide the user with the visual feedback or inspection, which we believe is a fruitful avenue for the future work.

References

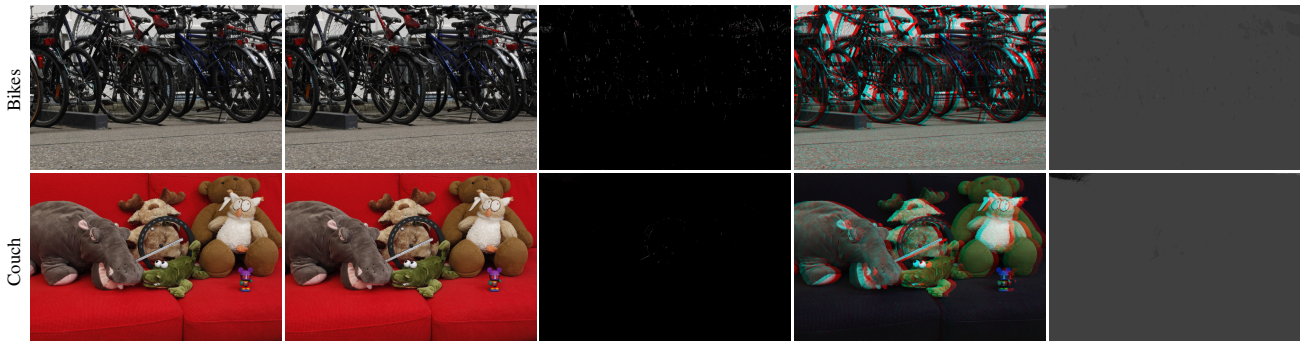
[1] A. Agrawal and R. Raskar. Gradient domain manipulation techniques in vision and graphics. In *ICCV Courses*, 2007. 6

[2] H. Bowles, K. Mitchell, R. W. Sumner, J. Moore, and M. H. Gross. Iterative image warping. *Comput. Graph. Forum*, 31(2):237–246, 2012. 1, 2

[3] T. F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal of Applied Mathematics*, 66(5):1632–1648, 2006. 4

[4] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang. Content-aware display adaptation and interactive editing for stereoscopic images. *IEEE Transactions on Multimedia*, 13(4):589–601, 2011. 1, 2

[5] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel. Adaptive image-space stereo view synthesis. In *VMV*,



(a) Reference image (b) Computed new view (c) Error (d) Anaglyph stereo (e) Labeling

Figure 8. *Linear disparity scaling.* (a) and (b) show the reference image and the computed new view, for which the depth at the reference view is linearly scaled by a factor of 10 was used as the target disparity map. (c) shows the error of the computed image against the ground truth, i.e., the 10th next image to the reference in the input light field (the darker the smaller error). (d) shows the anaglyph stereo image, and (e) shows the resulting labeling, where each step in the grayscale denotes an image index. The labeling should ideally look flat in this task.

pages 299–306, 2010. 1, 2

[6] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel. A perceptual model for disparity. *ACM Trans. Graph.*, 30(4):96, 2011. 1, 2

[7] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel. Apparent stereo: The cornsweet illusion can enhance perceived depth. In *IS&T/SPIE Symposium on Electronic Imaging*, pages 1–12, 2012.

[8] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, H.-P. Seidel, and W. Matusik. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph.*, page 184, 2012. 1, 2

[9] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, 1987. 3

[10] M. Grasmair and F. Lenzen. Anisotropic total variation filtering. *Applied Mathematics & Optimization*, 62(3):323–339, 2010. 3

[11] A. Handa, R. A. Newcombe, A. Angeli, and A. J. Davison. Applications of Legendre-Fenchel transformation to computer vision problems. Technical Report DTR11-7, Imperial College, Department of Computing, September 2011. 4

[12] S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik, and M. H. Gross. Computational stereo camera system with programmable control loop. *ACM Trans. Graph.*, 30(4):94, 2011. 2

[13] C. Kim, A. Hornung, S. Heinzle, W. Matusik, and M. H. Gross. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph.*, 30(6):190, 2011. 1, 3, 6, 7

[14] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. H. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, page 73, 2013. 5

[15] S. J. Koppal, C. L. Zitnick, M. F. Cohen, S. B. Kang, B. Ressler, and A. Colburn. A viewer-centric editor for 3d movies. *IEEE Computer Graphics and Applications*, 31(1):20–35, 2011. 2

[16] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. H. Gross. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.*, 29(4), 2010. 1, 2, 3

[17] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH*, pages 31–42, 1996. 2

[18] B. Masia, G. Wetzstein, C. Aliaga, R. Raskar, and D. Gutierrez. Display adaptive 3D content remapping. *Computers & Graphics*, 37(8):983–996, 2013. 2

[19] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, 37(8):1012 – 1038, 2013. 1, 2

[20] C. Olsson, M. Byrd, N. C. Overgaard, and F. Kahl. Extending continuous cuts: Anisotropic metrics and expansion moves. In *ICCV*, pages 405–412, 2009. 3

[21] T. Oskam, A. Hornung, H. Bowles, K. Mitchell, and M. H. Gross. OSCAM - optimized stereoscopic camera control for interactive 3D. *ACM Trans. Graph.*, 30(6):189, 2011. 2

[22] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 279–290, 2001. 1, 3

[23] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *ECCV*, pages 792–805, 2008. 4

[24] C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung. Megastereo: Constructing high-resolution stereo panoramas. In *CVPR*, pages 1256–1263, 2013. 3

[25] S. M. Seitz and J. Kim. The space of all stereo images. *International Journal of Computer Vision*, 48(1):21–38, 2002. 2

[26] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):11, 2011. 1

[27] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. H. Gross. StereoBrush: Interactive 2D to 3D conversion using discontinuous warps. In *SBM*, pages 47–54, 2011. 1, 2, 5

[28] C. Zach, M. Niethammer, and J.-M. Frahm. Continuous maximal flows and Wulff shapes: Application to mrfs. In *CVPR*, pages 1911–1918, 2009. 4

[29] M. Zwicker, W. Matusik, F. Durand, and H. Pfister. Antialiasing for automultiscopic 3D displays. In *EGSR*, pages 73–82, 2006. 5