Nonlinear Disparity Mapping for Stereoscopic 3D

Manuel Lang[1]   Alexander Hornung[1]   Oliver Wang[1]   Steven Poulakos[1,2]   Aljoscha Smolic[1]   Markus Gross[1,2]

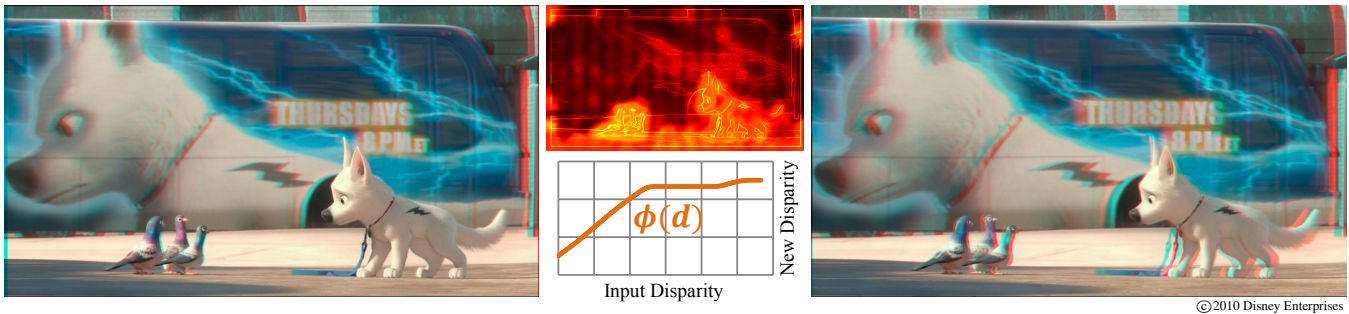[1]Disney Research Zurich   [2]ETH Zurich

Figure 1: *Our method retargets stereoscopic 3D video automatically to a novel disparity range, based on visual importance of scene elements and a nonlinear disparity mapping operator ϕ. This retargeting is accomplished using a novel stereoscopic image warping technique.*

Abstract

This paper addresses the problem of remapping the disparity range of stereoscopic images and video. Such operations are highly important for a variety of issues arising from the production, live broadcast, and consumption of 3D content. Our work is motivated by the observation that the displayed depth and the resulting 3D viewing experience are dictated by a complex combination of perceptual, technological, and artistic constraints. We first discuss the most important perceptual aspects of stereo vision and their implications for stereoscopic content creation. We then formalize these insights into a set of basic *disparity mapping operators*. These operators enable us to control and retarget the depth of a stereoscopic scene in a nonlinear and locally adaptive fashion. To implement our operators, we propose a new strategy based on *stereoscopic warping* of the input video streams. From a sparse set of stereo correspondences, our algorithm computes disparity and image-based saliency estimates, and uses them to compute a deformation of the input views so as to meet the target disparities. Our approach represents a practical solution for actual stereo production and display that does not require camera calibration, accurate dense depth maps, occlusion handling, or inpainting. We demonstrate the performance and versatility of our method using examples from live action post-production, 3D display size adaptation, and live broadcast. An additional user study and ground truth comparison further provide evidence for the quality and practical relevance of the presented work.

Keywords: Stereoscopy, 3D video, depth perception, disparity mapping, warping

1 Introduction

Stereoscopic 3D is on the cusp of becoming a mass consumer product. Cinemas show an increasing number of movies produced in 3D, TV channels are beginning to launch 3D broadcasts of sports events, and companies are offering 3DTV sets and Blu-ray 3D players. But despite these technological advances, the practical production of stereoscopic content that results in a natural and comfortable viewing experience in all scenarios is still a great challenge.

The fundamental problem lies in the complex interplay of human visual perception and the restrictions of display devices [Howard and Rogers 2002; Hoffman et al. 2008]. As a consequence, visual content must be adapted to the peculiarities of particular application scenarios. For fields other than stereoscopic 3D video, this content *retargeting* or *remapping* problem has been investigated extensively in computer graphics. For example, tone mapping techniques [Reinhard et al. 2005] exploit properties of our color perception to adapt HDR images to display devices of lower dynamic range and vice versa by nonlinear remapping of colors. Similarly, work on image and video retargeting [Shamir and Sorkine 2009] has shown how to perform spatially adaptive mapping of image content to different output formats by considering perceptual saliency cues.

In stereoscopic content production and display, similar remapping issues have to be addressed. Diverse studies and psychophysical experiments have revealed fundamental limitations of current stereoscopic display devices [Hoffman et al. 2008]. While today's 3D display technology can recreate the effect of vergence (vertical rotation of both eyes in opposite directions to maintain binocular vision), other important depth cues, such as accommodation (change of focus), cannot be faithfully reproduced as the resulting image is being displayed on a flat surface. This conflict has severe consequences; when displaying a close object on a distant screen, the strong negative disparity may result in an uncomfortable viewing experience and can cause temporary diplopia, the inability to fuse stereoscopic images. These effects are a major problem in practical 3D movie production. Content optimized for a standard 30 foot cinema screen will look completely different on a TV screen or a handheld display, and individual viewers can have vastly different viewing preferences. Hence, controlling and adapting disparity to the viewing situation is of central importance to the widespread adoption of stereoscopic 3D [Sun and Holliman 2009]. In addition, movie directors often employ (local) depth manipulation as an artistic and narrative device. All these issues have led to a complex set of best practice rules in the industry for how to film and display stereoscopic movies [Mendiburu 2009; Neuman 2009]. Implementing these rules requires considerable expertise on how to control disparity during filming and post-production.

A further significant problem is the realization of these guidelines in practice. Once stereo footage is recorded, it is no longer possible

to alter relevant parameters such as camera baseline or disparity range. In principle, techniques for image-based view interpolation [Zitnick et al. 2004] could be employed, but these methods tend to involve tasks such as estimating camera parameters, dense stereo reconstruction, and inpainting of occluded scene content. These are under-constrained and computationally complex problems, which cannot yet be solved with the necessary accuracy and robustness for general scenes and classical 2-view stereo footage. Therefore, movie and video producers have to resort to labor intensive and extremely costly manual editing of disparities (e.g., by compositing content from multiple stereo rigs of varying baseline). While this approach is expensive (but possible) in post-production for some scenarios, it is prohibitive for live broadcast where modifications of the disparity range have to be performed on the fly.

Our paper addresses the above mentioned problems with two main contributions. As a first contribution, we introduce *disparity mapping operators*. These operators are based on four central aspects of disparity in stereoscopy. We review these aspects from a perceptual point of view and discuss the resulting implications and requirements for stereoscopic content production and display. Our operators then formalize these insights and are the basis for a general framework for stereoscopic retargeting and disparity editing.

As a second contribution, we describe a conceptually simple but practical and powerful new technique for applying these disparity mapping operators to stereoscopic 3D footage. Our method is based on *stereoscopic image warping* instead of classical view interpolation. In contrast to previous works, our method requires only a sparse set of stereo correspondences which can be computed with sufficient robustness. We introduce novel disparity-based saliency measures and warp constraints which ensure consistent and content-adaptive remapping of the disparity range according to the chosen mapping operator. Additional support for manual disparity authoring seamlessly integrates this approach into existing production workflows. Using our warping approach, central problems of existing view interpolation methods such as camera calibration, accurate dense depth, and inpainting are avoided.

We demonstrate the versatility and practical relevance of our operators and warping technique on various types of stills and video. In particular, we present several applications of our method to central problems in stereo production: automatic disparity correction of live broadcast, nonlinear disparity editing and temporal disparity correction for movie post-production, retargeting of stereo footage to different display sizes, and 2D to 3D conversion of video. A user study is provided in order to validate our approach and the quality of our results and a ground truth comparison is used to analyze errors that arise from our stereoscopic image warping technique.

2 Related Work

Stereoscopic 3D production and display for movies or 3DTV is a challenging multi-disciplinary field, combining basic research on binocular vision and perception, camera and display technologies, as well as cinematography and art.

The capabilities of our visual system and depth perception have been the topic of numerous works and experiments in research on human vision [Burt and Juelsz 1980; Cutting and Vishton 1995; Howard and Rogers 2002]. One fundamental limitation is the range of disparities. As an example, we are unable to perceive extremely close and distant objects at the same time in 3D due to the large disparity range on our retina. Interestingly, however, our visual system still has quite strong abilities to compensate for inconsistent stereo cues, e.g., [Stelmach et al. 2000].

The rising popularity and recent developments of 3D display technology (e.g., [Matusik and Pfister 2004]) requires a reinvestigation of perceptual limitations in the context of the technological capabilities. Most of the current 3D display technology is based on display-ing a stereo image pair on a flat screen. This approach reproduces stereo cues such as vergence, but neglects other important depth-cues like accommodation. It has been shown that this discrepancy between accommodation and vergence yields problems such as distorted perception or visual fatigue [Hoffman et al. 2008; Lambooij et al. 2009], and considerable research efforts are invested to minimize these issues [Siegel and Nagata 2000; Akeley et al. 2004].

In stereoscopic content production, the most important tool to address such discrepancies between stereo cues is to adapt the range of disparities, i.e., the depth of a scene [Mendiburu 2009; Sun and Holliman 2009]. Besides pure adaption, however, control over scene depth is also an important artistic tool. Correspondingly there exists a complex set of cinematographic guidelines and rules on best practice in 3D movie making [Mendiburu 2009; Neuman 2009], as well as some prior work that allows for manually-driven disparity editing in specific application scenarios [Pritch et al. 2000; Feldmann et al. 2003; Wang and Sawchuk 2008]. However, a rigorous formalization of these principles for disparity editing under consideration of perceptual as well as production-related issues has not been achieved yet. Inspired by work on content retargeting and tone mapping [Reinhard et al. 2005; Weyrich et al. 2007; Shamir and Sorkine 2009] we present a solution for general *nonlinear disparity mapping operators* for stereoscopic 3D in Section 3.

Also on the technical level, disparity control of filmed stereoscopic video is a highly non-trivial problem, since novel views have to be generated that reflect the desired depth structure of the scene. The classical approach to this problem has been to perform image-based view interpolation, which either requires a very large number of densely sampled input images or additional accurate depth maps to achieve high quality results [Gortler et al. 1996; Levoy and Hanrahan 1996; Shade et al. 1998; Zitnick et al. 2004; Criminisi et al. 2007; Kim et al. 2008; Smolic et al. 2008; Bleyer et al. 2009]. One example of commercial software that uses image-based view interpolation for stereo editing is Ocula [the Foundry 2010]. These types of view interpolation involve a large number of computationally complex problems such as camera calibration, accurate depth, inpainting and rendering. Due to this complexity, fully automatic, sufficiently robust and accurate methods for cinematographic production and display adaptation are not available yet. There are also some techniques that provide a simplified manual interface for creating 3D scenes from video [van den Hengel et al. 2007], or generating stereographic sequences from single view input [Guttmann et al. 2009], but these methods require either calibration, static scenes, and manual tuning or dense depth estimation respectively. For small scale interpolation, image-based view morphing is an alternative [Seitz and Dyer 1996; Mahajan et al. 2009]. The great advantage of these methods is that they directly work in image space without the complex reconstruction and rendering. However, they are not suitable for general adaption of the scene's global depth, since they do not support the required nonlocal consistency constraints.

Recently, methods based on warping have shown to be powerful tools for complex operations on images and video which preserve the realism of the original input, including camera stabilization [Liu et al. 2009], optimizing image content [Carroll et al. 2009], and video retargeting [Krähenbühl et al. 2009; Wang et al. 2009]. Inspired by these works we present a novel technique for *stereoscopic image warping* in Section 4 which enables complex disparity editing of existing stereoscopic 3D footage.

3 Stereoscopic Disparity

As motivated in the introduction, stereoscopic 3D production and display is a complex field, involving a broad range of research and experience on human visual perception [Howard and Rogers 2002], display technology [Hoffman et al. 2008], and industrial best practice [Mendiburu 2009]. Addressing all involved issues requires multi-disciplinary research efforts where progress in one field might
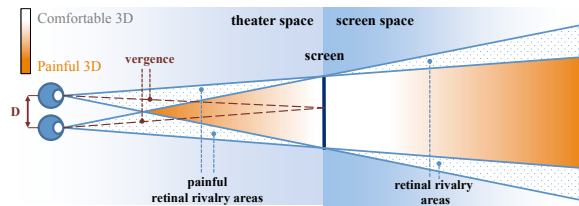
Figure 2: *Illustration of the stereoscopic comfort zone.*

lead to the need for new research in related fields. However, similar to the now well established fields of media retargeting or tone mapping, there exist a number of fundamental insights about stereoscopic perception and display, which are of highest relevance in application domains such as 3DTV and 3D cinema.

One of the central parameters in stereoscopy is *disparity*. This section is concerned with the discussion of four of the most important problems related to disparity [Howard and Rogers 2002; Mendiburu 2009]. We will provide an overview of these issues from a perceptual point of view, and describe how they are addressed in today's stereoscopic production pipeline in Section 3.1. Based on these basic rules and guidelines, we then propose a set of *disparity mapping operators* in Section 3.2 which formalize these ideas and provide a first basic and extendable framework for general disparity editing of stereoscopic 3D footage.

3.1 Background from Stereography

Disparity Range. Our visual system has several constraints regarding the admissible distance of corresponding points on the retina that still allows for proper depth perception. The central parameters influencing retinal disparity are the *interocular distance*, the *vergence* of the eyes, and the distance to the point of interest. For example, if we focus on a nearby object, the images of other objects in the distant background cannot be fused by our visual system anymore due to too large retinal disparities and will appear as double images (please refer to [Howard and Rogers 2002] for more detail). Depending on the above parameters, there is only a restricted disparity range around the *Horopter*, called Panum's area, which permits proper stereo vision and depth perception.

A central challenge in stereoscopic movie production is that current display technologies only have indirect control over these parameters, which they achieve by presenting a pair of slightly different images to the left and right eyes. The only parameter, which can be directly controlled, is the distance between corresponding features in the two displayed images, the *image disparity*. The actual retinal disparity then results from the convergence of the eyes, distance to and size of the screen, etc. In the following, when we discuss changing the "disparity" of a scene, we refer to the modification of the image disparity. A further problem is that important depth cues such as accommodation cannot be controlled at all; we have to focus our eyes on the screen surface, even if objects are positioned in front or behind the screen surface, resulting in conflicting depth cues. These technological limitations can lead to considerable problems, ranging from distortions of the 3D structure of a scene to visual fatigue [Hoffman et al. 2008]. With existing capture and display technologies these fundamental problems cannot be resolved. Hence, the disparity range for a displayed scene has to be adapted in order to minimize these issues.

In production, the admissible disparity range is the so called *comfort zone* (see Figure 2). Modifying stereoscopic content and disparity ranges to a generic comfort zone suitable for large population groups has been investigated, for example, in the context of Microstereopsis [Siegel and Nagata 2000]. A prominent solution today is *linear* disparity remapping [Sun and Holliman 2009; Mendiburu 2009]. Such a linear mapping changes the disparity in-

terval from a given range to the desired range. The introduced linear distortion of the disparity space amounts to uniform flattening of objects in the scene. Some concrete applications for linear disparity range mapping are the adaptation of stereoscopic content to display devices of different size, the uniform compression of scene depth for a more comfortable viewing experience, or moving scene content forward or backward by adding a constant offset to the disparity range. Practical values for disparity on a 30 foot cinema screen, are between +30 (appears behind screen) and -100 (appears in front of screen) pixels, assuming video with a width of 2048 pixels[Neuman 2009]. In practice such a mapping can be achieved by modifying the camera baseline (the interaxial distance) during filming, and by shifting the relative position of the left and right view after filming to control the absolute disparity offset. For instance, objects that are floating in front of the screen and intersect with the image borders will cause retinal rivalry (see Figure 2). In post-production this can be corrected using the floating window technique, which is a virtual shift of the screen plane towards the viewer [Mendiburu 2009]. In general, however, such adaptations have to be performed by expensive and cumbersome manual per-frame editing, since the camera baseline of recorded footage cannot be easily modified.

These disparity range limitations are the most obvious issue in stereoscopic perception and production. However, related to this limitation are a number of further issues, which we shall describe.

Disparity Sensitivity. Our ability to discriminate different depths decreases with increased viewing distance. One result from perceptual research is that the stereoacuity is inversely proportional to the square of the viewing distance [Howard and Rogers 2002]. This means that our depth perception is generally more sensitive and accurate with respect to nearby objects, while for distant objects other depth cues such as occlusion or motion parallax are more important [Banks et al. 2004].

This effect can be exploited in stereoscopic movie production by compressing the disparity values of distant objects. For example, a disadvantage of the previously mentioned linear range adaptation is that strong disparity range reduction leads to apparent flattening of objects in the foreground. Using the insights about stereoacuity, the decreased sensitivity to larger depths can be used to apply *nonlinear* remapping instead, resulting in less flattening of foreground objects. Effectively, this corresponds to a compression of the depth space at larger distances. This idea can be extended to composite nonlinear mapping, where the disparity range of single objects is stretched, while the space in between the objects is compressed. Such nonlinear operations which exploit the limitations in sensitivity of our visual system have been successfully employed in related areas such as media retargeting. But so far, they are difficult to apply to stereoscopic footage of live action, since this would require an adaptive modification of the camera baseline. In production, the only way to achieve such effects is complex multi-rigging by capturing a scene with camera rigs of varying baseline and manual composition in post-production [Neuman 2009; Mendiburu 2009].

Disparity Gradient. Besides limitations with respect to absolute disparity values, experiments have shown that our perception is subject to limits regarding the disparity gradients in an observed scene as well [Burt and Juelsz 1980]. In particular, the perception of different depth gradients strongly depends on local scene content, spatial relationships between objects, etc. Consequences range from distorted perception of the 3D structure of a scene to the inability to see proper stereo.

Exploiting the different types of gradient sensitivities of our visual perception (e.g., regarding color) has proven to be a valuable tool in research on tone mapping, where locally adaptive gradient domain processing is used for content-aware color remapping. In stereography, locally adaptive disparity modifications are important in two respects. On the one hand, it has to be ensured that the displayed gradients do not violate our perceptual limits. On the other, disparity gradient editing provides the possibility to redesign the depth
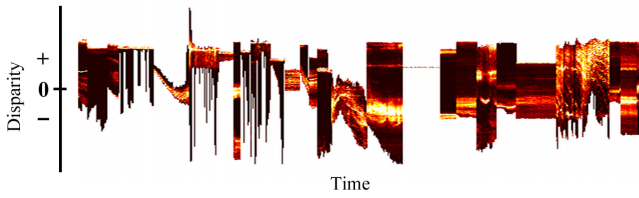
Figure 3: *A disparity storyboard for a 3D movie. In this plot, the range of estimated disparity values is plotted on the vertical axis on a frame-by-frame basis. The color indicates the frequency of the occurrence of disparity values. The velocity is visible in the changes in disparity histograms over time.*

structure of a scene on an object basis. This type of artistic freedom is a highly desired feature in post-production, but extremely difficult to achieve at the moment (e.g., using the previously mentioned multi-rigging) [Neuman 2009].

Disparity Velocity. The last important area is the temporal aspect of disparity. For real world scenes without conflicting stereo cues, it has been shown that our visual system can rapidly perceive and process stereoscopic information. The reaction time, however, can increase considerably for conflicting or ambiguous cues, such as inconsistent vergence and accommodation. Moreover, there is an upper limit to the temporal modulation frequency of disparity [Howard and Rogers 2002].

These temporal properties have considerable importance in the production of stereoscopic content. In the real world we are used to disparities varying smoothly over time. In stereoscopic movies, however, transitions and scene cuts are required. Due to the above mentioned limitations such strong discontinuities are perceptually uncomfortable and might again result in the inability to perceive depth [Mendiburu 2009]. Therefore, stereoscopic film makers often employ a continuous modification and adaption of the depth range at scene cuts in order to provide smooth disparity velocities, so that the salient scene elements are at similar depths over the transition. Additionally, such depth discontinuities can be exploited explicitly as a storytelling element or visual effect and are an important tool used to evoke emotional response [Neuman 2009]. Figure 3 illustrates disparity histograms of a 3D movie over time, where such smooth transitions are visible. These disparity storyboards are an important part of the planning required for 3D productions.

We summarize the four central aspects of disparity which we utilize to design the disparity mapping operators in the following section:

Disparity Range: Mapping of the global range of disparities, e.g., for display adaptation.
Disparity Sensitivity: Disparity mapping for global or locally adaptive depth compression and expansion.
Disparity Gradient: Content-adaptive disparity remapping by modifying disparity gradients.
Disparity Velocity: Temporal interpolation or "smoothing" between different disparity ranges at scene transitions.

These operations are of essential importance for the generation and display of 3D footage, be it during post-production of movies or real-time correction of live broadcasts. In the following section we will formalize these insights into corresponding disparity mapping operators and then present a novel framework that allows us to perform complex disparity editing on existing stereo footage.

3.2 Disparity Mapping Operators

We will first consider disparity mapping operators on a conceptual level. Section 5.1 will then provide examples for relevant application scenarios. Without loss of generality we assume that the input footage is recorded with a stereo camera rig or is approximately rec-

tified. For a digital stereo image pair (I_l, I_r) let $\mathbf{x} \in \mathbb{R}^2$ be a pixel position in the left image I_l. We define the disparity $d(\mathbf{x}) \in \mathbb{R}$ as the distance (measured in pixels) to the corresponding pixel in I_r (and vice versa). The range of disparities between the two images is an interval $\Omega = [d_{min}, d_{max}] \subset \mathbb{R}$. Our disparity mapping operators will be defined as functions $\phi : \Omega \to \Omega'$ which implement the rules and guidelines described in the previous section by mapping an original range Ω to a new range Ω'. For illustration we refer to the examples in Section 5.1.

Linear Operator: Globally linear adaptation of a disparity $d \in \Omega$ to a target range $\Omega' = [d'_{min}, d'_{max}]$ can be obtained by a mapping function

$$\phi_l(d) = \frac{d'_{max} - d'_{min}}{d_{max} - d_{min}}(d - d_{min}) + d'_{min}. \quad (1)$$

By changing the interval width of Ω', the depth range can be scaled and offset such that it matches the overall available depth budget of the comfort zone (e.g., Section 5.1 and Figure 11).

Nonlinear Operator: Global nonlinear disparity compression can be achieved by any nonlinear function, e.g.,

$$\phi_n(d) = \log(1 + sd), \quad (2)$$

with a suitable scale factor s. For more complex, locally adaptive nonlinear editing, the overall mapping function can be composed from basic operators. For example, given a set of different target ranges $\Omega_1, \ldots, \Omega_n$ and corresponding functions $\phi_0, \ldots, \phi_n$, the target operator would be:

$$\phi_a(d) = \begin{cases} \phi_0(d), & d \in \Omega_0 \\ \ldots & \ldots \\ \phi_n(d), & d \in \Omega_n \end{cases}. \quad (3)$$

An elegant approach to generate such complex nonlinear functions in a depth authoring system is to either use the histogram of disparity values (as shown in Figures 3 and 4) for identifying dominant depth regions, or to analyze the visual saliency of scene content in image space. These so called saliency maps $S(\mathbf{x}) \in [0, 1]$ (see Figure 5) represent the level of visual importance of each pixel and can be generated either automatically by the system (see also Section 4.2) or manually by the user. From the saliency map, the algorithm can infer which disparity ranges Ω_i are occupied by important objects, and which regions are less important. From these importance values, which essentially correspond to the first derivative ϕ'_a, the actual disparity operator can be generated as the integral $\phi_a(d) = \int_0^d \phi'_a(x)dx$ (please see Figures 1 and 7 for examples).

Gradient Domain Operator: In addition to local adaptivity in disparity space as in ϕ_a, the remapping of disparity gradients allows for additional spatial adaptivity in image space. Retargeting operators for disparity gradients with spatial adaptivity can be defined based on visual importance maps $S(\mathbf{x})$ as functions $\phi_\nabla(\nabla d(\mathbf{x}), S(\mathbf{x}))$. An example for adaptive compression using interpolation between a linear and a nonlinear map ϕ_l and ϕ_n is

$$\phi_\nabla(\nabla d(\mathbf{x}), S(\mathbf{x})) = S(\mathbf{x})\phi_l(\nabla d(x)) + (1 - S(\mathbf{x}))\phi_n(\nabla d(x)). \quad (4)$$

The actual disparity mapping operator can then be reconstructed from ϕ_∇ using methods from gradient domain processing [Agrawal and Raskar 2007].

Temporal Operator: Temporal adaptation and smoothing, as it is required for smooth scene transitions or visual effects, can be defined by weighted interpolation of two or more of the previously introduced operators, e.g.,
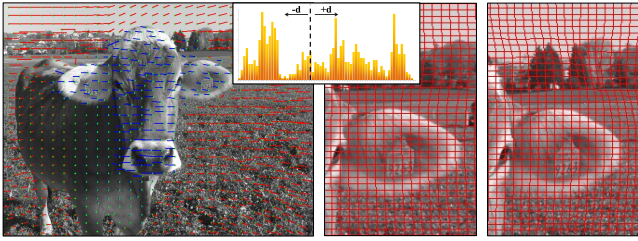
$$\phi_t(d, t) = \sum_i w_i(t)\phi_i(d), \quad (5)$$

Figure 4: *Left: stereo correspondences with color coded dispari-*
ties (red positive, blue negative) and the disparity histogram for the
cow example after pruning. Right: close-ups of the warped stereo
pair showing the deformed isolines with respect to the input views.

where $w_i(t)$ is a suitable weighting function. An example for tem-
poral interpolation and the resulting disparity histograms is given
in Section 5.1, Figure 9.

These different operators in Eq. (1)-(5) provide the basic function-
ality required to implement the set of central disparity operations
presented in Section 3.1. In the following section, we will present
our novel image-based stereoscopic warping scheme for applying
these operators to stereo footage. Section 5.1 will then provide con-
crete applications for these operators in production.

4 Stereoscopic Warping

We now know that disparities have to be adapted to the stereoscopic
comfort zone due to our fundamental perceptual and technical lim-
itations. However, once these conflicts are sufficiently minimized,
our depth perception is quite robust, even if the resulting 3D scene
is not geometrically consistent. This is due to the additional depth
information from cues such as relative size and order of objects in
a scene, and motion parallax [Cutting and Vishton 1995; Siegel and
Nagata 2000]. In the following section we exploit these properties
and present an algorithm for automatic disparity remapping based
on stereoscopic warping.

The basic idea is to first compute a set $\mathcal{F}$ of sparse feature cor-
respondences $(\mathbf{x}, \mathbf{x}')$ between the left and right view of a stereo
image pair (I_l, I_r) (Section 4.1). We then compute a novel, image-
and disparity-based visual saliency map S (see also Section 3.2),
which measures the visual importance of each pixel in the spatial
and in the depth domain (Section 4.2). Using our disparity map-
ping operators ϕ, the correspondences F, and the saliency map S,
we can compute a stereoscopic warp of the stereo pair, such that
the resulting output views fulfill the desired disparity constraints
defined by ϕ (Section 4.3).

4.1 Sparse Stereo Correspondences

Sparse feature correspondences between the two images (I_l, I_r)
can be estimated robustly using well established standard tech-
niques [Baker and Matthews 2004; Lowe 2004], hence we refer to
these works for detail on the basic correspondence matching. Op-
tionally we exploit downsampled dense correspondence informa-
tion [Werlberger et al. 2009] between I_l and I_r for large textureless
image regions which are too ambiguous for sparse feature match-
ing. Outliers can be removed automatically [Sattler et al. 2009].

Depending on scene content the resulting feature set $\mathcal{F}$ generally
has an irregularly clustered distribution of correspondences. More-
over, many features are not temporally stable over a longer video
sequence, but disappear after a few frames. Since our warping al-
gorithm requires only a sparse set of features, we apply a spatially
anisotropic pruning algorithm to $\mathcal{F}$ which favors temporally sta-
ble correspondences and is adaptive to depth discontinuities. Cor-
respondences are first sorted by their lifetime, so that long living
pairs receive a high priority and correspondences which appeared

only for a couple of frames receive a low priority. We then apply
a greedy procedure to remove low priority correspondences around
those with a high priority. Let $(\mathbf{x}_l, \mathbf{x}_r) \in \mathcal{F}$ be a high priority
correspondence pair with disparity $d(\mathbf{x}_l)$. Our pruning algorithm
removes all pairs $(\mathbf{x}_l', \mathbf{x}_r') \in \mathcal{F}$ with

$$\left\| \begin{pmatrix} \mathbf{x}_l \\ d(\mathbf{x}_l) \end{pmatrix} - \begin{pmatrix} \mathbf{x}_l' \\ d(\mathbf{x}_l') \end{pmatrix} \right\| < r. \tag{6}$$

This isotropic distance measure in image and disparity space results
in a locally adaptive anisotropic filter in image space only (similar
to the idea of the Fast Bilateral Filter [Paris and Durand 2006]).
Pruning is performed symmetrically for the positions $\mathbf{x}_r$ as well.
In principle, the radius r depends on the image resolution and the
disparity range. However, the results of our warping algorithm are
quite insensitive to different feature densities so that we could sim-
ply use a value of $r = 10$ in our experiments.

Figure 4 shows an example of the resulting features and the cor-
responding disparity histogram. This algorithm combines the re-
spective strengths of different methods for feature estimation and
provides a robust way to automatically compute a sparse but suffi-
ciently accurate set $\mathcal{F}$ of correspondences between stereo pairs.

4.2 Depth and Image Saliency

In order to determine which parts of the input images can be dis-
torted by our warp without creating visible artifacts, we need a vi-
sual importance map S for the stereo image pair. Our approach
to compute S is twofold. First, we use image-based importance
measures which are able to capture the coarse and fine scale de-
tails of image content, such as prevalent edges or textured regions.
In addition, we have the sparse disparity information from the pre-
viously computed stereo correspondences. This allows us to ex-
ploit the depth dimension as an additional source of information
to estimate visual saliency. Accordingly, we compute a composite
saliency map as a weighted combination

$$S(\mathbf{x}) = \lambda S_i(\mathbf{x}) + (1 - \lambda)S_d(\mathbf{x}), \tag{7}$$

for all pixels $\mathbf{x} \in I_l$ where S_i represents the image-based saliency
and S_d our disparity-based saliency. S_i is generated from the sum
of a local edge map and the global scale method of Guo et al. [2008]
(see Figure 5) for each stereo channel individually.

The disparity saliency map S_d can be computed by any operator on
the range of disparities of correspondences in $\mathcal{F}$. A simple but ef-
fective solution is to assume that foreground objects generally catch
our visual attention more than the background of a scene, which is
a reasonable assumption for many application scenarios (see Sec-
tion 3.1). So for a correspondence set $\mathcal{F}$ comprising a disparity
range $\Omega = [d_{min}, d_{max}]$, we assign high saliency values to dispar-
ities close to d_{min} and a low saliency to disparities close to d_{max}.
Saliency values are then interpolated over the non-feature pixels
(see Figure 5). Note that in principle more complex disparity-based
saliency estimators are possible (see also Section 5.1).

Figure 5 shows all components of the saliency computation. Dark
areas in the final map are parts of the scene that are more likely
to be distorted by the warp to accommodate movement within the
images. For weighting S_i and S_d our current implementation uses
a value $\lambda = 0.5$.

4.3 Warping

Our aim is now to warp the stereo image pair (I_l, I_r) such that the
range of disparities Ω of the stereo correspondences $\mathcal{F}$ is mapped
to a new range defined by a disparity mapping operator $\phi : \Omega \to
\Omega'$. This means we have to compute a pair of warping functions
(w_l, w_r) which map coordinates from (I_l, I_r) to a pair of output
images (O_l, O_r), respectively, i.e., $O_l \circ w_l = I_l$ and $O_r \circ w_r = I_r$,

Figure 5: *Individual saliency components and final automatically generated saliency map for a scene. From left to right: left image of a stereo pair, local edge saliency, global texture saliency, disparity-based saliency, combined saliency map S.*

subject to $d(O_l, O_r) = \phi(d(I_l, I_r))$. Note that in principle warping only a single image would be sufficient. However, by distributing the required deformation to both images, we are more flexible regarding the admissible disparity mapping operations without introducing noticeable visual artifacts. To compute these warps we employ the same basic methodology as existing warp-based methods for video retargeting [Shamir and Sorkine 2009]: we define a set of constraints on the functions (w_l, w_r) which can then be solved as a nonlinear least-squares energy minimization problem.

Stereoscopic Constraints. The most central set of constraints applies the disparity mapping operator ϕ to the stereo correspondences $(\mathbf{x}_l, \mathbf{x}_r) \in \mathcal{F}$. For each correspondence pair we require

$$w_l(\mathbf{x}_l) - w_r(\mathbf{x}_r) - \phi(d(\mathbf{x}_l)) = 0, \tag{8}$$

meaning that the disparity of a warped correspondence pair $(w_l(\mathbf{x}_l), w_r(\mathbf{x}_l))$ should be identical to applying the disparity mapping operator ϕ to the original disparity $d(\mathbf{x}_l)$.

Since the above constraints only prescribe relative positions, we require a small set of absolute position constraints which fix the global location of the warped images. We compute these position constraints for the 20% temporally most stable feature correspondences, i.e., those features which have been detected throughout a sequence of frames in the video. The warped positions are defined by the average previous position and the novel disparity:

$$w_l(\mathbf{x}_l) = \frac{\mathbf{x}_l + \mathbf{x}_r}{2} + \frac{\phi(d(\mathbf{x}_l))}{2}$$
$$w_r(\mathbf{x}_r) = \frac{\mathbf{x}_l + \mathbf{x}_r}{2} - \frac{\phi(d(\mathbf{x}_l))}{2} \tag{9}$$

Eq. (8) and (9) define the basic stereoscopic warping constraints so that the warped images match the target disparity range Ω'.

Temporal Constraints. For video with moving scene elements one has to ensure that local image distortion is properly transferred along the local motion flow [Werlberger et al. 2009] between successive video frames. The local image distortion can be measured based on the derivatives of the warp. Let $\partial w_x^t / \partial x$ denote the partial derivative of the x-component of the warp w^t at time t, and let $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$ be two corresponding pixels in I_{t-1} and I_t, respectively. The transfer of the warp distortion is then expressed by

$$\frac{\partial w_x^t}{\partial x}(\mathbf{x}_t) = \frac{\partial w_x^{t-1}}{\partial x}(\mathbf{x}_{t-1}). \tag{10}$$

This constraint is enforced for the y-component $\partial w_y^t / \partial y$ as well and performed for the left and the right image warp independently.

Saliency Constraints. Besides these novel stereoscopic and temporal warp constraints, we additionally employ a set of standard constraints which minimize the perceivable visual distortion. The idea is to enforce a certain rigidity of the warp in salient regions, and to allow for larger image distortions in non-salient regions. Hence, the constraints consist of terms for

- Distortions: $\frac{\partial w_x}{\partial x} = \frac{\partial w_y}{\partial y} = 1$,
- Bending of edges: $\frac{\partial w_x}{\partial y} = \frac{\partial w_y}{\partial x} = 0$,
- Overlaps: $\frac{\partial w_x}{\partial x} \wedge \frac{\partial w_y}{\partial y} > 0$.
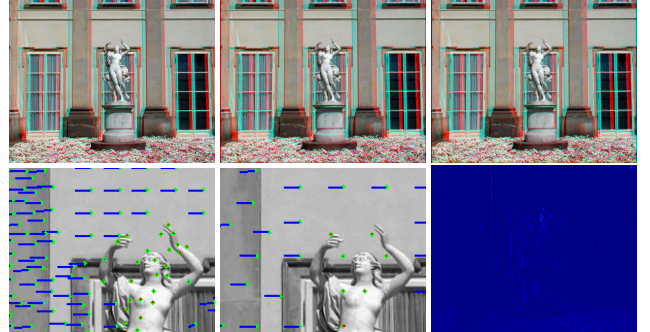

Figure 6: *Warping example showing stability over different numbers of stereo correspondences. Upper row, left to right: original stereo input, disparity mapping results (depth range increased) computed with about 2000 features, the same result using only about 200 features. Bottom row: close-ups showing the respective feature density and a difference image of the warped images.*

During the actual warp computation these constraints are then weighted by the saliency map S in Eq. (7) to achieve an adaption of the warp to the image-content. Since these basic constraints are identical to related work on video retargeting, please refer for example to [Krähenbühl et al. 2009] for details.

Implementation. Our implementation of these warping constraints follows the standard procedure in image-based warping: The constraints are converted into energy terms so that the computation of the warps (w_l, w_r) can be solved as an iterative nonlinear least-squares problem. In our current implementation we simply sum all of the above energy terms and weight the saliency constraints by multiplication with the saliency map S. The warp-induced deformation is illustrated in Figure 4 by overlaying isolines of the original input images. Figure 6 is an example with differing number of stereo correspondences and shows that the results of our stereoscopic warping are quite insensitive to the number of features.

In addition to automatic constraints it would be interesting to include the possibility to manually add high level constraints regarding region positions or global lines [Krähenbühl et al. 2009]. Since at its core our warp is similar to previous warping methods, the inclusion of these techniques is straightforward. However, as we show in our results, the current automatic solution already provides very acceptable results for a variety of stereoscopic 3D footage.

5 Results

As motivated in Section 3, the question of how the disparity range should be adapted for different types of stereoscopic footage depends strongly on the particular target application. Our goal was to achieve practical disparity retargeting that can be employed in actual application scenarios. Hence, we present nonlinearly and linearly mapped results for three important application scenarios. We also evaluate the quality of our method quantitatively by a ground truth comparison, and present the results of a user study to validate the perceptual quality of the warping.
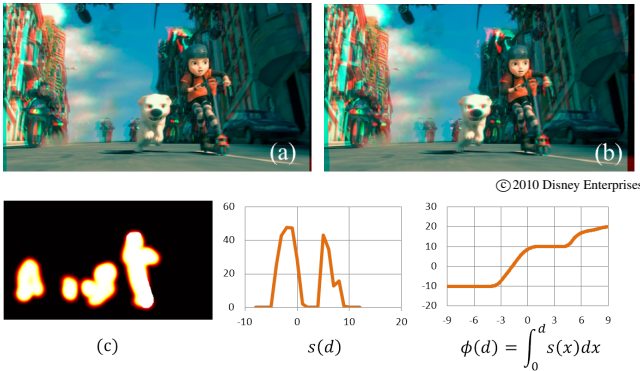
$s(d)$

$\phi(d) = \int_0^d s(x)\,dx$

(c)

Figure 7: *Post-Production. For an input frame (a) and a given importance map (c) the average importance for individual disparities $s(d)$ can be computed. This is automatically converted into a nonlinear depth mapping operator $\phi(d)$ as described in Section 3.2. The resulting image in which the key characters are emphasized by stronger depth is shown in (b).*
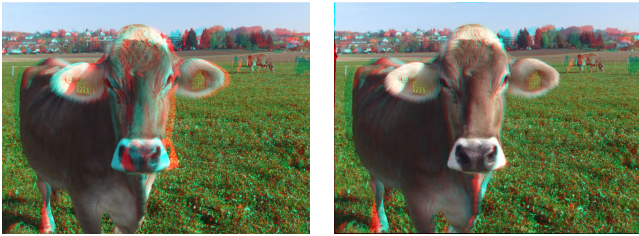

Figure 8: *Nonlinear disparity remapping. The disparity range of the original (left) is quite large leading to diplopia on large screens. Our nonlinearly remapped image (right) displays the cow behind the screen and compresses the depth range without apparent flattening of the cow's head.*

The results in this section are presented as red (left) - cyan (right) anaglyph images (optimized for zoomed on-screen display). More results and stereoscopic videos are included in the supplemental material. A free stereoscopic player is available at [3dtv.at 2010]. Note that in the anaglyph images, changes in disparity can generally also be estimated from the different displacement of the red and cyan channels (e.g., Figure 8).

5.1 Applications

The three production scenarios we present in this section include nonlinear and linear editing for post-production, automatic disparity correction, and display adaptation. Furthermore we illustrate the versatility of our method with an example for 2D to 3D conversion.

Post-Production. The first major application area of our disparity mapping operators and warping is post-production of stereoscopic content. We may assume a studio environment where skilled operators apply software tools within an interactive workflow to edit previously captured material. Using the proposed methodology and algorithms, depth composition can be modified and authored by combining different nonlinear and linear disparity operators.

Examples and results for nonlinear and linear disparity editing are shown in Figures 1, 7, 8, and 9. In Figure 1 we modify the global scene depth structure with a nonlinear function which emphasizes foreground content and compresses empty space in-between while retaining the maximum disparity of the background. In Figure 7 we exploit a visual saliency map to automatically design an adaptive disparity mapping operator. It enhances salient regions and simultaneously compresses the depth of unimportant regions (see also Eq. (3)). In Figure 8 the images were captured with a consumer 3D
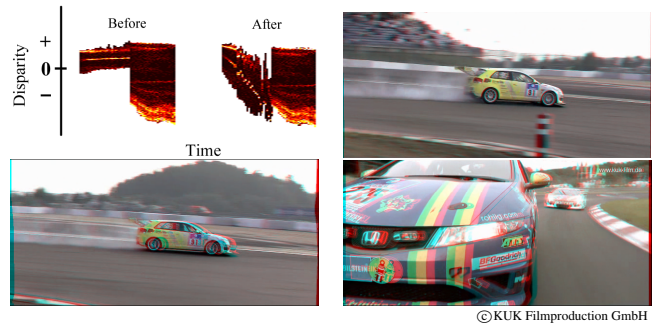

Figure 9: *Temporal adaptation for a scene cut. Before adaptation both sequences have considerably different disparity histograms with a clearly visible discontinuity in-between (upper left). Our temporal disparity mapping operator adapts the disparity range of the first sequence (first frame of the sequence in the upper right) to the disparity range of the second sequence (bottom row) and thus achieves a smoother transition.*

camera (Fuji Finepix 3D). The resulting disparity range in combination with negative parallax is too large for a comfortable viewing experience on large screens. We compressed the depth nonlinearly by moving the cow behind the screen surface (positive parallax) and in addition applied a discontinuous nonlinear map retaining the dimensionality of the cow's head without altering the maximum disparity. A further example is shown in Figure 6.

Another important scenario in stereoscopic content production relates to the temporal adaptation of the depth structure in scene transitions and the focus on salient objects therein (Section 3.1). Currently, cinematographers are designing depth storyboards in advance and modification of convergence by global image shift is the only tool to smooth over shot transitions. Our methods enable us to compensate for the sharp disparity jumps by slowly adapting the disparity ranges of the previous and/or current scenes. Figure 9 depicts the disparity histograms before and after correction.

Automatic Disparity Correction. Stereography in live action content production is a difficult art. Camera parameters such as baseline and vergence have to be adjusted carefully to ensure a high-quality view experience while keeping the overall action within the stereoscopic comfort zone. Settings are adjusted to match a certain depth range in which the action is expected to take place. In movie productions such decisions are taken by the directors, can be adjusted as appropriate, tested, and shots are repeated if necessary.

This is not possible in live broadcast scenarios or for the amateur home user. Any error will immediately lead to degradation in viewing quality or even result in diplopia. 3D sports broadcast is a popular and timely example. Movements of camera and objects are fast, spontaneous, often unpredictable, and interleaved with rapid scene cuts. This frequently leads to violations of the stereoscopic window or transgressions of admissible disparity ranges. Similar considerations apply to stereoscopic footage captured by amateurs. Simple shift convergence for correction will not help if the overall depth range of the scene is sufficiently large. Instead, careful limitation and compression of the disparity is required. Figure 10 displays examples for automatic disparity correction of such content.

Display Adaptation. A third application area is 3D display adaptation and retargeting. It is motivated by the observation that 3D content optimized for certain target display size and viewing distance (e.g., theatrical) will appear differently on a different medium (e.g., 3DTV). In order to retain a high viewing quality and the artistic intention, disparity adaptation is necessary when reformatting 3D content, e.g., from theatrical to TV or even to a handheld device. Examples of depth editing are illustrated in Figure 11.

2D to 3D Conversion. In order to demonstrate the versatility of our method we illustrate an example for 2D to 3D conversion.

Figure 10: *Automatic correction of disparity. The original stereo pairs are shown on the left, our result is on the right. The cropped racing car captured with strong negative disparities and a large overall scene depth results in the so-called framing problem. With a simple linear disparity scaling the car is pushed behind the screen without increasing the background disparity. In the bottom example, taken with a consumer 3D camera, the subject was moving towards the camera and finally exceeded the maximum disparity range. In our result, the global disparity range has been adapted so that the background remains at constant depth while the foreground is pushed closer to the screen.*

Recreating 3D stereo pairs from existing 2D images or video involves an expensive and cumbersome interactive workflow. Recently, Guttmann et al. [2009] presented a novel approach that simplifies this task by solving for dense depth maps from sparse user scribbles to generate stereographic sequences. Using our method the requirement for dense depth can relaxed, so that the sparse scribbles alone are already sufficient. The warp interpolates the pixel disparities and generates a stereographic image pair from a single input image (see Figure 12).

5.2 Ground Truth Comparison

The purpose of the ground truth experiment was to assess the visual quality of the our warping approach. We utilized a publicly available data-set generated with a multi-camera rig [Mobile 3DTV 2010]. We picked two views (numbers 8 and 10) from 3 different data-sets and then applied the warp to generate intermediate and extrapolated views (numbers 7, 9, and 11). The extrapolated views represent a doubling of the camera baseline while the interpolated view corresponds to a baseline of 0. We then compared the quality of our result to the known ground truth images using both a perceptually motivated structural similarity metric SSIM [Wang et al. 2004] and by computing the RMSE of the difference images. Single pixel shifts can cause a high RMSE error while contributing little to perceived image quality. Hence, perceptual measures like SSIM are generally better suited for such types of comparisons. The values for extrapolated views in Table 1 are averaged over the two views 7 and 11. As Table 1 and the images in Figure 13 reveal, our method is able to compute interpolated and extrapolated views of a scene that are perceptually indistinguishable from the original, provided the disparity range is not too large. Please note, however, that geometrically consistent view interpolation has not been an explicit goal of this work.

5.3 User Study

To further assess the suitability of *warping* for disparity mapping we conducted a user study with 22 subjects and 15 test cases. The goals of the user study were to show specifically that (1) warping

Figure 11: *Display adaptation into both directions. The middle images are the original stereo pairs, while the left images feature a linear reduction of the disparity range to 50% and the right images an increase to 200%. These results show that our method allows to preserve the initial depth structure relative to the screen geometry to adapt content to actual viewing conditions.*

Figure 12: *2D to 3D conversion. From a single 2D input image and providing only a sparse set of disparity cues shown as rough scribbles, our method produces a convincing 3D result.*

Figure 13: *Ground truth comparison showing a known ground truth image (left) versus our warped image (middle left), and absolute difference images of warped and known ground truth for both baseline reduction (middle right) and extension (right).*

indeed results in a perceivable change of a scene's depth structure, and that (2) the quality is not degenerated by visual artifacts.

We performed the study with a line-wise polarized 46 inch full HD display manufactured by Miracube. All 22 participants were tested for their ability to perceive depth on a stereo screen using a random dot stereogram. One subject had to be excluded due to a negative test. The remaining 21 subjects took part in a pairwise evaluation [David 1963] of video material composed from short clips from 3D Hollywood movies, 3D sports and other professional stereo video. Every side-by-side comparison featured the original scene as well as a disparity mapped version of it using our method. For nine test cases we doubled the initial disparities and for six test cases we reduced them by a factor of 0.5. We randomized the order of videos as well as the locations of the original and manipulated videos. For every comparison the participants had to answer the following two questions:

1: Q: Which video features more depth? A: *Left* or *Right*

2: Q: Which video is the original? A: *Left*, *Right*, or *Don't know*

Depth perception. In total we received 311 valid votes regarding the depth impression. Overall, 253 votes (81%) correctly recognized the example with larger disparity range. Kendall's coefficient of agreement [David 1963], which measures the interobserver variability for pairwise comparison tests, is $u = 0.391$ with a p-value < 0.01. Two sequences (street view, train) only reached a correspondence of 66%. Both videos feature a very strong perspective depth cue. One sequence had a recognition rate of only 55%. This sequence contains fast and complex motion cues for scene elements confirming well-known observations in stereo perception, such as

Table 1: *SSIM and RMSE values of ground truth comparison with 3 different datasets for view interpolation and extrapolation. SSIM=1 means no difference to the original,*

Dataset	1 interp.	1 extrap.	2 interp.	2 extrap.	3 interp.	3 extrap.
SSIM	.9999	.9998	.9999	.9999	.9999	.9999
RMSE	.0313	.0334	.0235	.0242	.0228	. 0236

the domination of motion parallax cues and the resulting difficulty to focus on binocular depth effects. 17 participants correctly recognized the depth mapping for 11 or more sequences. One subject drastically diverges with only 5 recognitions (9 is the next better result). Removing these outliers (three sequences and one subject) from the evaluation leads a recognition rate of 88%.

Quality. Question 2 was answered in 56% cases as *Don't know*. 25% of the votes correctly identified the original, but 19% were wrong and assumed that the manipulated video was actually the original video. Some originals were correctly recognized by over 80%, but some of the manipulated sequences were also considered originals by over 80%.

The major conclusion we draw from the study is that disparity mapping based on image warping can change the depth structure of a scene in a perceptually believable way without introducing distracting visual artifacts. As we have demonstrated earlier, the suitability of a particular operator (local, global, linear, nonlinear) highly depends on the application, artistic intention, scene content and motion, and other criteria.
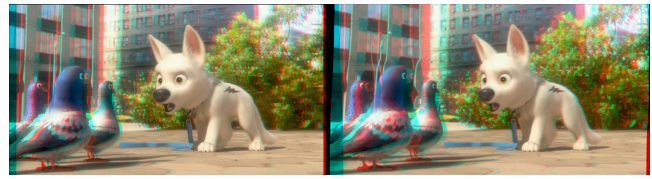
5.4 Limitations

A limitation of our method is depicted in Figure 14. For image regions in which the disparity changes rapidly, such as around the pigeons' heads, the sparse features and warp can lead to visible distortions. These artifacts could be addressed by using a higher feature count, denser depth information, or by adding manual constraints to enforce feature preservation. Such methods have been successfully proposed in work on image warping [Krähenbühl et al. 2009].

However, it is very interesting to observe that, when viewed in 3D, such artifacts are often visually less apparent due to the complex compensation mechanisms of our visual system. These phenomena clearly deserve additional research in the context of stereoscopic warping. From research on media retargeting it is also well known that there are certain limits for warping-based methods on how strongly image content may be deformed before artifacts become visible [Shamir and Sorkine 2009]. But since most often the required deformation for disparity adaptation lies in the range of only 1-2% of the overall pixel resolution, these limits are typically not reached. In none of the examples presented in this paper did we observe visible artifacts, which is further proved by the user study. Finally, our method is limited in the extent to which we can modify the camera baseline for nearby objects, since such operations imply explicit handling of occluded areas to avoid conflicting cues.

6 Conclusions

In this paper we presented a set of *disparity mapping operators* providing a basic formalization of perceptually motivated and production-oriented rules and guidelines for nonlinear disparity editing of stereoscopic 3D content. In order to implement these operators we proposed a novel technique based on *stereoscopic warping*, which allows us to deform input video streams in order to meet a desired disparity range. We applied our techniques to three different scenarios, all of which are of very high practical relevance in stereoscopic production and display, and demonstrated that automatic image-based warping could be used as a general alternative for rendering even complex depth manipulations. The quality of

© 2010 Disney Enterprises

Figure 14: *Limitations. For image regions with frequent and strong changes in disparity, the sparse features and the warp can lead to distortions visible around the pigeons' heads. Interestingly, however, such distortions seem to be compensated to some extent by our visual system during stereoscopic viewing.*

stereoscopic warping was evaluated with a ground truth experiment and a user study.

Automatic disparity correction could be implemented in future generation stereo camera systems to support the cinematographer or cameraman in realtime. In addition to professional live broadcast systems, consumer electronic systems could also benefit from such methods, since amateur users generally do not have the experience and background required for proper stereo capture. Moreover, algorithms for disparity adaptation could be implemented as part of future 3D display devices or TVs enabling viewers to control disparity on-the-fly in order to match the display size and the user preferences, much like we control aspect ratio, contrast, or color today. In future research we would like to refine our nonlinear mapping operators to accommodate additional features of stereoscopic perception. The current user study was limited to assess the suitability of warping for disparity mapping. Future studies will investigate the influence of various nonlinear and local operators on the perceived quality of the results. Furthermore, we want to investigate to what extent conflicting cues, such as inaccurate occlusions, can be compensated for by additional cues like motion parallax.

References

3DTV.AT, 2010. Stereoscopic player, Jan. http://www.3dtv.at/.

AGRAWAL, A., AND RASKAR, R. 2007. Gradient domain manipulation techniques in vision and graphics. In *ICCV Courses*.

AKELEY, K., WATT, S. J., GIRSHICK, A. R., AND BANKS, M. S. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. 23*, 3, 804–813.

BAKER, S., AND MATTHEWS, I. 2004. Lucas-Kanade 20 years on: A unifying framework. *IJCV 56*, 3, 221–255.

BANKS, M. S., GEPSHTEIN, S., AND LANDY, M. S. 2004. Why is spatial stereoresolution so low? *Journal of Neuroscience 24*, 2077–2089.

BLEYER, M., GELAUTZ, M., ROTHER, C., AND RHEMANN, C. 2009. A stereo approach that handles the matting problem via image warping. In *CVPR*, 501–508.

BURT, P., AND JUELSZ, B. 1980. A disparity gradient limit for binocular fusion. *Science 208*, 4444 (5), 615–617.

CARROLL, R., AGRAWALA, M., AND AGARWALA, A. 2009. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph. 28*, 3.

CRIMINISI, A., BLAKE, A., ROTHER, C., SHOTTON, J., AND TORR, P. H. 2007. Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. *Int. J. Comput. Vision 71*, 1, 89–110.

CUTTING, J. E., AND VISHTON, P. M. 1995. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Handbook of perception and cognition, Perception of space and motion*, W. Epstein and S. Rogers, Eds., vol. 5. Academic Press, San Diego, CA.

DAVID, H. A. 1963. *The Method of Paired Comparisons*. Charles Griffin & Company.

FELDMANN, I., SCHREER, O., AND KAUFF, P. 2003. Nonlinear depth scaling for immersive video applications. WIAMIS.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *SIGGRAPH*, 43–54.

GUO, C., MA, Q., AND ZHANG, L. 2008. Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform. *CVPR*.

GUTTMANN, M., WOLF, L., AND COHEN-OR, D. 2009. Semi-automatic stereo extraction from video footage. In *ICCV*.

HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3 (3), 1–30.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth*. Oxford University Press, New York, USA.

KIM, M.-B., LEE, S., CHOI, C., UM, G.-M., HUR, N.-H., AND KIM, J.-W. 2008. Depth scaling of multiview images for auto-multiscopic 3D monitors. In *3DTV08*.

KRÄHENBÜHL, P., LANG, M., HORNUNG, A., AND GROSS, M. 2009. A system for retargeting of streaming video. *ACM Trans. Graph. 28*, 5.

LAMBOOIJ, M., IJSSELSTEIJN, W., FORTUIN, M., AND HEYNDERICKX, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology 53*, 3, 030201.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *SIGGRAPH*, 31–42.

LIU, F., GLEICHER, M., JIN, H., AND AGARWALA, A. 2009. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph. 28*, 3.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2, 91–110.

MAHAJAN, D., HUANG, F.-C., MATUSIK, W., RAMAMOORTHI, R., AND BELHUMEUR, P. N. 2009. Moving gradients: a path-based method for plausible image interpolation. *ACM Trans. Graph. 28*, 3.

MATUSIK, W., AND PFISTER, H. 2004. 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Trans. Graph. 23*, 3, 814–824.

MENDIBURU, B. 2009. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press.

MOBILE 3DTV, 2010. Stereo video data-sets, Jan. http://sp.cs.tut.fi/mobile3dtv/stereo-video/.

NEUMAN, R., 2009. Personal Communication with Robert Neuman, Chief Stereographer, Disney Animation Studios.

PARIS, S., AND DURAND, F. 2006. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV (4)*, 568–580.

PRITCH, Y., BEN-EZRA, M., AND PELEG, S. 2000. Automatic disparity control in stereo panoramas (omnistereo). In *OMNIVIS*.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.

SATTLER, T., LEIBE, B., AND KOBBELT, L. 2009. SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter. In *ICCV*.

SEITZ, S., AND DYER, C. 1996. View morphing. In *SIGGRAPH 96*, 21–30.

SHADE, J., GORTLER, S. J., LI-WEI, H., AND SZELISKI, R. 1998. Layered depth images. In *SIGGRAPH*, 231–242.

SHAMIR, A., AND SORKINE, O. 2009. Visual media retargeting. In *SIGGRAPH ASIA Courses*.

SIEGEL, M., AND NAGATA, S. 2000. Just enough reality: Comfortable 3-D viewing via microstereopsis. *IEEE Transactions on Circuits and Systems for Video Technology 10*, 3 (4), 387–396.

SMOLIC, A., MLLER, K., DIX, K., MERKLE, P., KAUFF, P., AND WIEGAND, T. 2008. Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems. In *ICIP*, IEEE, 2448–2451.

STELMACH, L. B., TAM, W. J., MEEGAN, D. V., AND VINCENT, A. 2000. Stereo image quality: effects of mixed spatio-temporal resolution. *IEEE Transactions on Circuits and Systems for Video Technology 10*, 2, 188–193.

SUN, G., AND HOLLIMAN, N. 2009. Evaluating methods for controlling depth perception in stereoscopic cinematography. *Stereoscopic Displays and Virtual Reality Systems XX, Proceedings of SPIE 7237* (1).

THE FOUNDRY, 2010. Ocular, Nuke, Jan. http://www.thefoundry.co.uk/.

VAN DEN HENGEL, A., DICK, A. R., THORMÄHLEN, T., WARD, B., AND TORR, P. H. S. 2007. Videotrace: rapid interactive scene modelling from video. *ACM Trans. Graph. 26*, 3, 86.

WANG, C., AND SAWCHUK, A. A. 2008. Disparity manipulation for stereo images and video. SPIE, vol. 6803.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4, 600–612.

WANG, Y.-S., FU, H., SORKINE, O., LEE, T.-Y., AND SEIDEL, H.-P. 2009. Motion-aware temporal coherence for video resizing. *ACM Trans. Graph. 28*, 5.

WERLBERGER, M., TROBIN, W., POCK, T., WEDEL, A., CREMERS, D., AND BISCHOF, H. 2009. Anisotropic Huber-L1 optical flow. In *British Machine Vision Conference (BMVC)*.

WEYRICH, T., DENG, J., BARNES, C., RUSINKIEWICZ, S., AND FINKELSTEIN, A. 2007. Digital bas-relief from 3D scenes. *ACM Trans. Graph. 26*, 3, 32.

ZITNICK, C. L., KANG, S. B., UYTTENDAELE, M., WINDER, S. A. J., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. *ACM Trans. Graph. 23*, 3, 600–608.