Vis Comput (2013) 29:685–694
DOI 10.1007/s00371-013-0804-6

ORIGINAL ARTICLE

Attention-Aware Disparity Control in interactive environments

Ufuk Celikcan · Gokcen Cimen · E. Bengu Kevinc · Tolga Capin

Published online: 26 April 2013

Abstract Our paper introduces a novel approach for controlling stereo camera parameters in interactive 3D environments in a way that specifically addresses the interplay of binocular depth perception and saliency of scene contents. Our proposed Dynamic Attention-Aware Disparity Control (DADC) method produces depth-rich stereo rendering that improves viewer comfort through joint optimization of stereo parameters. While constructing the optimization model, we consider the importance of scene elements, as well as their distance to the camera and the locus of attention on the display. Our method also optimizes the depth effect of a given scene by considering the individual user's stereoscopic disparity range and comfortable viewing experience by controlling accommodation/convergence conflict. We validate our method in a formal user study that also reveals the advantages, such as superior quality and practical relevance, of considering our method.

Keywords Stereoscopic 3D · Disparity control · Interactive 3D · User attention · Real-time graphics · Accommodation/convergence conflict

U. Celikcan (✉) · G. Cimen · E.B. Kevinc · T. Capin
Bilkent Universitesi Bilgisayar Muhendisligi Bolumu,
Bilkent 06800 Ankara, Turkey
e-mail: celikcan@acm.org

G. Cimen
e-mail: gokcen.cimen@cs.bilkent.edu.tr

E.B. Kevinc
e-mail: kevinc@cs.bilkent.edu.tr

T. Capin
e-mail: tcapin@cs.bilkent.edu.tr

1 Introduction

Recent advances in stereoscopic displays and 3D TVs, 3D digital cinema, and 3D enabled applications have increased the importance of stereoscopic content creation and processing. However, several challenges remain in providing realistic but comfortable viewing experience to users with stereoscopic products. One of the principal challenges is a need for applying the underlying principle of 3D perception of the human visual system and its capabilities/limitations for displaying content in stereoscopic displays.

Binocular viewing of a scene is created from two slightly different images of the scene in the two eyes. These views are produced by stereoscopic rendering parameters, which are camera separation and convergence distance of cameras. The difference in the views, or screen disparities, create a perceived depth around the display screen. The main concern of stereoscopic 3D content creation is determining the comfortable range of this perceived depth, also called *the comfort zone*.

Recent research has made progress in controlling the perceived depth range, mostly in post production pipeline [3, 12, 19]. On the other hand, different from offline production, in an interactive environment where the position of the camera is dynamically changing based on the user input, there is a need for a control system to keep the perceived depth in the comfortable target range. Examples for such controllers are the work of Lang et al. [12] for post-production disparity range adjustment and the work of Oscam et al. [16] for real-time disparity range adaptation.

An example for an interactive setting is a game environment where the stereoscopic output changes dynamically. For such an environment, finding optimized stereoscopic camera parameters, i.e., camera convergence distance and interaxial separation to retarget dynamic scene depth

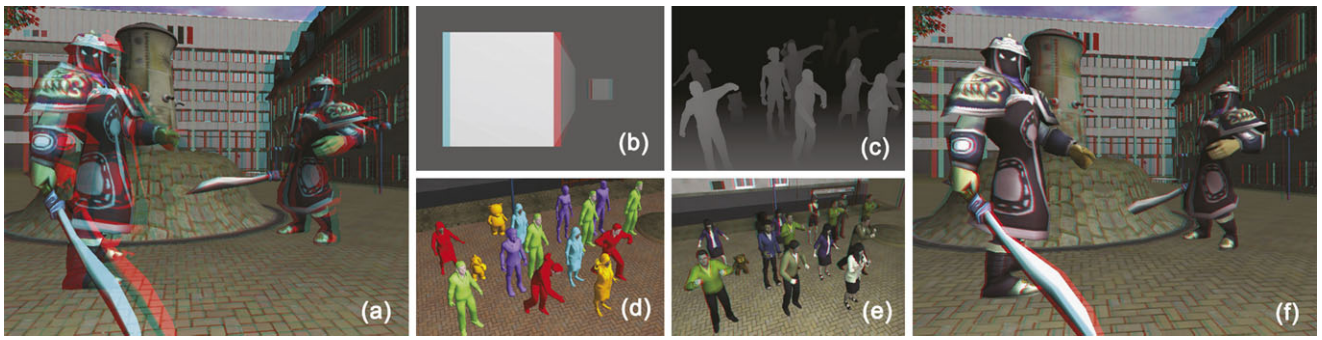


Fig. 1 (**a**) An example capture of the scene with Naive method. (**b**) Disparity limit calibration. (**c**) Depth map of a captured scene. (**d**) Significance score coloring of scene elements. (**e**) Output stereoscopic image with DADC. (**f**) Capture of the scene with DADC

to comfortable target depth range brings a great challenge. Even though previous works manage to control and limit the perceived depth to comfort zone of the users, there is also a need to define parameters for preventing the violation of accommodation/convergence conflict. This conflict can cause severe consequences in such interactive stereoscopic environments in long-term use. The inability of fusion, also called diplopia, is one of the major problems that emerge because of accommodation/convergence conflict, and further problems include eye-strain, visual fatigue and even headache after prolonged exposure.

In this work (Fig. 1), we aim to address the challenges of presenting a comfortable viewing experience to users in an interactive scene, by controlling and limiting target depth range to the comfort zone and eliminating accommodation/convergence violations as much as possible. For mapping scene depth to the specific depth range, our method automatically finds optimized stereo camera parameters in real-time. In order to avoid accommodation/convergence conflict, we consider the distribution and importance of scene elements. For this purpose, the convergence plane is moved so that significant elements are shown with relatively sharper focus. This motivation comes from that the location of the convergence plane, on which scene elements are captured with exactly zero disparity, should tend to be nearer to elements with higher significance during the search, assuming each element of interest in the scene content carries a significance score that is assigned by the content creator.

2 Related work

With the recent advances in stereoscopic systems, the focus on stereoscopic camera control has gained momentum and a number of techniques have been proposed for stereoscopic post-production pipeline and editing of stereoscopic images.

3D camera systems and stereo acquisition The conventional way for capturing real scenes is with two physical camera equipments. One of the recent approaches which focus on production of high quality stereoscopic content capture is presented by Zilly et al. [21]. This system analyzes the captured scene by two real cameras and specifies the proper camera calibration parameters. Heinzle et al. [6] focus on controlling the rig directly, with a control loop that consists of capture and analysis of 3D stereoscopic parameters.

Stereoscopic editing on still images Recent work on stereoscopic image editing focuses on correction of imperfect stereoscopic images and videos. Koppal et al. [11] present an editor for live stereoscopic shots. They concentrate on the viewer's experience and propose modifying camera parameters in the post processing as well as previewing steps. Lang et al. [12] present a nonlinear disparity mapping method in order to retarget the depth range in the produced stereoscopic images and videos to different displays and viewing conditions. Didyk et al. [2] have also recently proposed a disparity model that estimates the perceived disparity change in processed stereoscopic images, and perform psychophysical experiments in order to derive a metric for modeling disparity. Didyk et al. [3] also proposed an extended luminance-contrast aware disparity model, and presented disparity retargeting as one of its applications.

Stereo parameter adjustment in virtual environments Post processing and image shifting methods are used for retargeting disparity in offline applications such as digital cinema and 3D content retargeting. On the other hand, interactive applications require real-time techniques. Among recent works, the geometrical framework to map a specified depth range to the perceived depth range is described by Jones et al. [10]. Their method is proposed for generating still images, but it can also be used for virtual scenes. Oskam et al. [16] present a controller for finding camera convergence and interaxial separation, which gives a final disparity value for the viewed frame. These parameters change automatically by taking minimum and maximum scene depth values

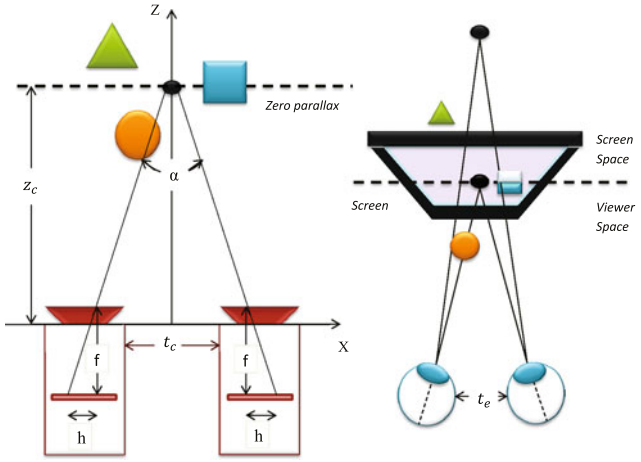


Fig. 2 A virtual camera setup with parallel sensor-shift (left) and the corresponding reconstruction of stereoscopic 3D scene

into account in order to handle excessive binocular disparities which are generated because of unpredictable viewer motion.

3 Background

As our system makes use of the characteristics of binocular vision and stereo geometry, in this section we summarize the basic principles behind them.

Depth perception Depth cues, which help the human visual system to perceive spatial relationships between objects, constitute the core part of depth perception. These visual cues can be categorized as pictorial, oculomotor, binocular, and motion-related cues [7]. Pictorial cues, such as occlusion, shadow, shading, relative size, relative height, texture gradient, are extracted from a single and flat 2D view; whereas oculomotor depth cues represent depth perception that is obtained through eye movements. Motion parallax, motion perspective, and kinetic depth are the motion-based depth cues. The two types of binocular depth cues are named as convergence and retinal disparity, which are covered in detail in the following.

Stereo geometry The binocular depth cue makes use of the fact that left and right eyes view the world from slightly different angles, which results in slightly different retinal images, forming binocular vision. The parameters that are used in the human visual system by their real world correspondences are *binocular disparity* and *vergence*. Binocular disparity represents the difference between the two eyes; whereas vergence arises due to eye movements and allow fixating at a point of interest.

In stereoscopic image creation, the main difficulty arises while controlling the stereoscopic camera parameters. There

Table 1 The review of the perceptual effects of stereo parameters (adapted from Milgram and Kruger [15])

		Disparity	Perceived Depth	Object Size
t_c	Increases	*Increases*	*Increases*	*Constant*
	Decreases	*Decreases*	*Decreases*	*Constant*
Z_c	Increases	*Decreases*	*Shifts Forward*	*Constant*
	Decreases	*Increases*	*Shifts Backward*	*Constant*

are two principal parameters for disparity: *interaxial separation* (t_c) and *convergence distance* (Z_c), as illustrated in Fig. 2. While convergence distance corresponds to the distance between the camera and the plane in focus, the interaxial separation corresponds to the separation between the two cameras. The camera separation, or interaxial separation (t_c) directly affects the disparity and eventually the amount of depth perceived in the final image. The convergence distance, on the other hand, does not affect the overall perceived depth, but increasing the convergence distance decreases the screen parallax. Table 1 summarizes the perceptual effects of the stereoscopic camera parameters.

Given the parallel camera geometry in Fig. 2, the image disparity of an object with scene distance Z depends on interaxial separation (t_c) and convergence distance (Z_c), and is given as:

$$d = f t_c \left(\frac{1}{Z_c} - \frac{1}{Z} \right). \tag{1}$$

In this equation, f denotes the focal length of the cameras. The conversion from image disparity d to screen parallax p simply requires scaling the image disparity from image sensor metric to display size metric, by multiplying it with a scale factor W_s / W_i, where W_i and W_s denote the image sensor width and screen width, respectively.

$$p = d(W_s / W_i). \tag{2}$$

While maintaining stereoscopic depth, the viewer reconstructs a point for each object on and around the screen. The reconstructed depth Z_r of this point, while the viewer is observing from a physical distance Z_w, is given as

$$Z_r = \frac{Z_w t_e}{t_e - p} = \frac{Z_w t_e}{t_e - d(W_s / W_i)}, \tag{3}$$

where t_e is the human interocular distance, for which the physiological average is approximately 65 mm.

The convergence distance gives the distance where the two cameras converge; and on the plane at that distance the retinal positions of objects appear at the same point which results in objects appearing at the physical screen surface ($Z = Z_c$). This condition is called *zero parallax setting*.

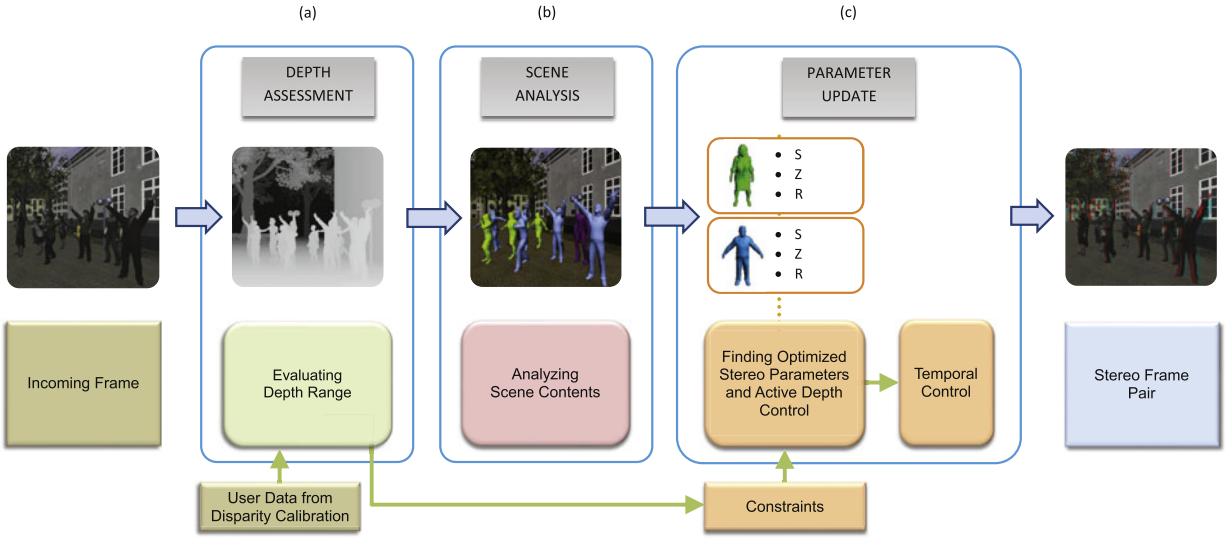


Fig. 3 Overview of the main phase of our approach. (**a**) In the first stage, visible scene depth extrema information is gathered. This information in combination with the data collected from the disparity calibration phase is fed into the optimization as system constraints. (**b**) The scene content analysis stage, as outlined in Algorithm 1, extracts $\{S, Z, R\}$ information of significant elements in the visible scene. (**c**) The system searches for the optimal parameter set $\{Z_c, t_c\}$ seeking to keep significant scene elements inside the comfort zone while maximizing the perceived depth feeling. The system output is finalized by applying *temporal control* to the optimization output

Two conditions occur when object distances Z are different from Z_c. In the first case, $(Z > Z_c)$, the object appears inside the screen space, which is viewed behind the display screen. When this condition occurs, the object has a positive disparity, or screen parallax. On the other hand, in the case $(Z < Z_c)$, the object has a negative disparity, or parallax. These objects appear as if they are physically located in front of the screen.

Physiological experiments have proven that the human visual system has more tolerance to positive parallax than negative parallax [14]. However, it is still restricted to comfortably perceive all objects which appear in positive or negative parallax regions. It has been shown that locating the scene in a limited area around the screen surface gives more reasonable results for avoiding accommodation/convergence conflicts.

Accommodation/convergence conflict The conclusion pointed out by several earlier studies [20] on the issue of stereoscopic comfort zone is that the amount of perceived depth in stereoscopic displays should be limited; and the conflicts related to accommodation and convergence should be controlled. The accommodation/convergence conflict happens for all planostereoscopic displays, i.e. displays where the views are presented on a planar screen. This conflict is caused by the fact that when looking at the stereoscopic 3D display, viewer's eyes converge on the reconstructed depth Z_r, while they are forced to focus on the display plane. This is in contrast to natural vision in the real world, where the human visual system operates such that the eyes converge and accommodate at the same point.

4 Approach

Our approach consists of a calibration phase and a main phase. In the calibration phase, the depth perception range of the user is obtained interactively. Perceived depth range is changeable in light of user's personal stereoscopic comfort limits. For this purpose, the user designates the personal disparity extrema, so that the disparity is not too high in order to avoid eye-straining visual artifacts like diplopia, or too low resulting in low depth feeling. This calibration stage is needed to be performed only once per user, before starting the interactive stage.

During the main phase (Fig. 3), for the incoming frame, we first analyze the depth range of the scene from the given view position. Consecutively, we perform an analysis of the scene contents, in terms of their layout under the given viewing condition. For this purpose, for each object in the view, we consider its significance score, its distance to the camera and center of display, and construct an optimization problem that we solve to calculate the stereo parameters, t_c and Z_c. Our method also makes use of temporal coherency constraint, so that the stereo parameters change smoothly between frames.

4.1 Depth Range Control (DRC)

Our method is an extension of the methods that control the depth range in a given scene. Among which, the most widely used one is Depth Range Control (DRC) method and our approach includes this method as a special case. Therefore, we first explain DRC, before discussing our approach in detail.

Algorithm 1 Scene content analysis algorithm

```
e[ ] ← getSignificantElements()
    ▷ Acquiring all significance score assigned elements in the current scene
j ← 0
for ∀e[i] do
    if e[i] is visible in the current frame then
        e[i].Z ← ForwardDistanceFromCamera()
        if e[i].Z ≤ D_max then
                ▷ D_max: maximum forward distance allowed
            o[j] ← e[i]
                ▷ implies o[j].S ← e[i].S and o[j].Z ← e[i].Z
            o[j].R ← RadialDistanceFromCameraAxis()
            j ← j + 1
        end if
    end if
end for
return o[ ]
```

It is possible to approximate the perceived disparity by geometrically modeling the stereoscopic vision with respect to a given depth-range which may be adjusted by the viewer. According to this approach, interaxial separation and convergence distance can be formulated [20] by using similar triangles in the stereo vision geometry. This, for an image-shift camera convergence setup, results in:

$$Z_c = \frac{Z_{\max} Z_{\min} (d_{\max} - d_{\min})}{(Z_{\max} d_{\max} - Z_{\min} d_{\min})}, \tag{4}$$

$$t_c = \frac{Z_{\max} Z_{\min} (d_{\max} - d_{\min})}{f (Z_{\max} - Z_{\min})}, \tag{5}$$

where $Z_{\max}$: The distance between the camera and the farthest object in the virtual world, $Z_{\min}$: The distance between the camera and the nearest object in the virtual world, $d_{\max}$: Maximum disparity, i.e., the positive disparity of the farthest object, $d_{\min}$: Minimum disparity, i.e., the negative disparity of the nearest object.

Jones et al. [10] applied this model to adjust the target depth range of still images only. Guttmann et al. [5] used the model for recreating stereographic sequences from 2D input by estimating the correct target depth distribution and optimizing the target disparity map. Oskam et al. [16] developed a similar method for interactive applications for optimizing stereo rendering parameters with respect to control points each assigned a certain desired depth. In the special case with only two constraints, one for each depth extremum, their system simplifies to Eq. (4) and Eq. (5) above.

In any case, the mentioned methods are based on mapping the depth range, without consideration of the distribution of the objects in the scene. Therefore, we believe that employing DRC method alone is not sufficient in enhancing the perceived stereo vision effect, as psychological elements directly affect the creation of stereo vision, especially in interactive applications. In this regard, we develop an attention-aware system which involves real-time analysis of scene contents as well as depth range assessment for user-specific disparity control.

4.2 Dynamic Attention-Aware Disparity Control (DADC)

As overviewed in the previous section, it is known that objects which are located in the 3D comfort zone of the user are easier to be observed. Thus, significant scene elements that draw user's attention should be located closer to this region. However, in a pre-produced interactive scene, it is necessary to move the convergence plane instead, placing it as near as possible to the region that attracts the user's attention the most, while maintaining the total disparity of the scene as high as possible and not violating the user's disparity range.

With this goal in mind, the main phase of our stereoscopic 3D control system is composed of the following three consecutive stages.

4.2.1 Depth range calculation

Since the maximum and the minimum distances observed by the virtual camera have a direct effect on screen disparity and thus the depth experienced by the user, we need to gather visible scene depth extrema information. This is achieved by a number of min-max reduction passes on the depth buffer [4]. The system runs this normally costly procedure in real-time (i.e., within the allowed per-frame time budget) by efficient utilization of the GPU.

This information in combination with the data collected from disparity calibration of the user is fed into the optimization as system constraints, and is also used in the two special non-optimization cases, as explained in detail later.

4.2.2 Analysis of scene contents

Having adopted interactive environments as our main consideration, we make the following arguments in conjunction with our objective function that is explained in the next section:

- The user navigates towards scene elements that attract his attention more.
- The user tends to have significant scene elements centered in his view.

Based on these assumptions, we evaluate the overall significance of a scene element with respect to the three criteria below:

S: significance score of the element.
Z: forward distance of the element from camera.
R: radial distance of the element from forward camera axis.

Here, we assume that scene elements had been assigned significance scores by the content creator that would appropriately predict the user's relative attention towards them such that e.g., in a first-person game environment the autonomous enemies should have been assigned higher scores compared to other scene elements.

Our scene content analysis algorithm progresses as outlined in Algorithm 1.

4.2.3 *Optimization of stereo parameters with active depth control*

For establishing our objective function to be optimized, we first formulate an energy term $E_o(Z_c, t_c)$ that penalizes the distance of the convergence plane from scene elements with relatively higher significance score and/or with relatively lower radial distance from the user's center of attention.

In order to minimize visual artifacts like ghosting associated with significant scene elements, the higher the significance score of an element the closer convergence plane should move towards it through minimization of $E_o(Z_c)$ thus keeping that element in relatively sharper focus.

Several methods have been proposed for computational modeling of visual attention [8]. Studies have converged on a two-component framework for attention; where viewers selectively direct their attention in an image, to objects in a scene using both (i) bottom-up, image-based saliency cues and (ii) top-down, task-dependent cues.

For precise detection of the center of attention, a perceptually based system should include some sort of eye-tracking technology as it deals with the extent of features across the user's retina or at least head-tracking technology that mimics eye-tracking by the observation that resting eye gaze can approximately track head orientation. However, when no eye or head tracking exists, as is the case with most stereoscopic viewing settings, we are to conform to the assumption [17] that the user always looks toward the center of the display device. Considering this, by minimizing $E_o(Z_c)$, the resulting convergence plane should also move closer towards scene elements with relatively less radial distance from the forward axis of virtual camera i.e., display center.

Following this line of thought, $E_o(Z_c)$ is formulated as

$$E_o(Z_c) = \sum_{i=1}^{n} \frac{S_i}{R_i^2}(Z_i - Z_c)^2, \tag{6}$$

where n is the number of significant scene elements found in the scene analysis stage.

We use a second energy term $E_d(Z_c, t_c)$ which pursues to maximize total scene disparity and, therefore, total perceived depth. Formulation of $E_d(Z_c, t_c)$ follows the regular disparity calculation (Eq. (1)) s.t.

$$E_d(Z_c, t_c) = \sum_{i=1}^{n} S_i f t_c \left(\frac{1}{Z_c} - \frac{1}{Z_i}\right), \tag{7}$$

hence aggregating weighted disparity associated with each significance assigned scene element. Here, disparities are also weighted with respective significance scores S_i.

We construct the objective function as the total energy function $E(Z_c, t_c)$ s.t.

$$E(Z_c, t_c) = \hat{E}_o(Z_c) - \hat{E}_d(Z_c, t_c), \tag{8}$$

Here $\hat{E}_o(Z_c)$ and $\hat{E}_d(Z_c, t_c)$ are the normalized energies s.t.

$$\hat{E}_o(Z_c) = E_o(Z_c)/(Z_{\max} - Z_{\min})^2, \tag{9}$$

$$\hat{E}_d(Z_c, t_c) = E_d(Z_c, t_c)/(d_{\max} - d_{\min}). \tag{10}$$

This way with appropriate normalization, the need to express $E(Z_c, t_c)$ as a weighted sum of $E_o(Z_c)$ and $E_d(Z_c, t_c)$ with weights that are to be fine-tuned for every different setting and every different user is avoided.

Consequently, by minimizing $E(Z_c, t_c)$, the system searches for the optimal parameter set by mediating the minimization of $E_o(Z_c)$ with the maximization of $E_d(Z_c, t_c)$, thus seeking to keep significant scene elements inside the comfort zone while maximizing the perceived depth feeling.

The system minimizes $E(Z_c, t_c)$ subject to constraints:

$$d_{\max} \geq f t_c \left(\frac{1}{Z_c} - \frac{1}{Z_i}\right) \geq d_{\min}, \quad \forall i \,|\, 1 \leq i \leq n, \tag{11}$$

with $d_{\max}$ and $d_{\min}$ obtained from disparity calibration phase. The constraints ensure that during the optimization scene depth is actively mapped into the perceivable depth range of the user as initially determined.

The nonlinear system is globally optimized within the parameter space by improved stochastic ranking-based evolutionary strategy (ISRES) algorithm [18]. The ISRES algorithm, a major representative of the state of the art in constrained optimization, is based on a simple evolution strategy augmented with a stochastic ranking that decides by carrying out a comparison, which utilizes either the function value or the constraint violation. With the incorporation of ISRES implementation in NLopt library [9] using modern multi-core processor technology via multi-threading, we achieve optimization at interactive speed so that the system is able to produce the updated stereo parameters continually as e.g., the user navigates through a scene.

Frames with only a single element of interest When the system finds a single significance assigned element visible, it places the element at the screen i.e., $Z = Z_c$ and computes interaxial separation using the DRC method.

Frames without an element of interest For frames containing no significance assigned element, our system switches to complete DRC mode and computes the stereo parameters accordingly.

Temporal control Stereoscopic 3D rendering parameters are recalculated for each frame as a desired solution. On the other hand this situation may cause undesired visual artifacts if changes in parameters occurring between consecutive frames are considerably high or happening more frequently than tolerable. In order to uphold temporal coherence, the system produces the final parameter set for the processed frame by passing each newly computed parameter through a threshold function $f(\cdot)$ s.t.

$$f\big(x(t)\big) = \begin{cases} x(t-1) + x_1, & \text{if } x(t) - x(t-1) \leq x_1; \\ x(t-1) + x_2, & \text{if } x(t) - x(t-1) \geq x_2; \\ x(t-1) + k(x(t) - x(t-1)), & \text{otherwise.} \end{cases} \tag{12}$$

where $x_1 \in \mathbb{R}^-$, $x_2 \in \mathbb{R}^+$ and k is chosen to be $0 < k < 1$.

5 Experimental evaluation

To evaluate our method, we tested it in two different scenes in pair-wise comparisons to the DRC only approach and the Naive approach. The Naive approach uses fixed stereo parameters that are initialized with DRC method at the beginning of each test session.

5.1 Subjects

We recruited 15 subjects, with a mean age of 25. The subjects were among voluntary undergraduate and graduate students with computer science background; and most of them did not have previous detailed experience on rendering on stereoscopic displays. Prior to the study, each subject candidate was tested for proper stereoscopic visual acuity using random dot stereogram test and those who failed the test did not participate in the user study. The subjects were not informed about the purpose of the experiment.

5.2 Equipment

We used a 2.20 GHz Quad-Core laptop with 6 GB RAM for rendering; and a 40 inch 3D display with active shutter glasses, with a resolution of 1920×1080. The subjects were seated at a viewing distance of 2 m.

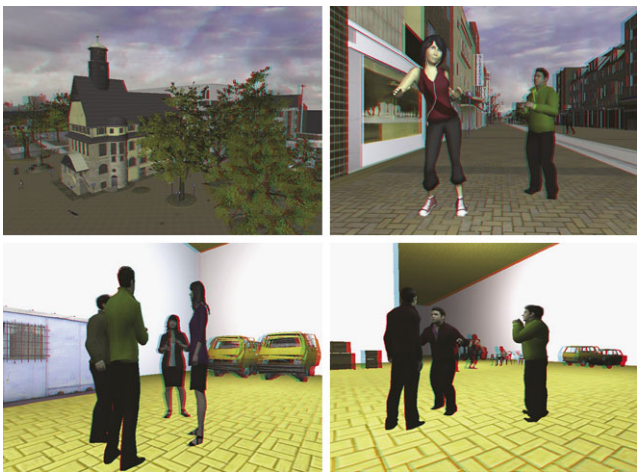


Fig. 4 First row shows snapshots of outdoor scene, second row shows of indoor scene

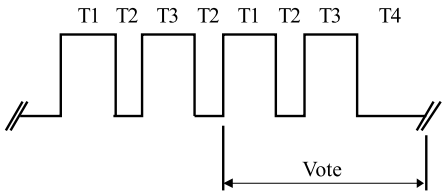


Fig. 5 Presentation of test material

5.3 Scenes

We built two interactive scenes (Fig. 4) for the tests. The first scene contains an indoor setting, where several groups of human characters, each of which performing various gestural movements, randomly distributed in a room. The second one contains an urban outdoor setting that presents a more dynamic environment in terms of variety of characters and their actions, as well. Virtual characters were assigned relatively higher significance in both scenes. In each test, the user was asked to navigate freely in the environment.

5.4 Procedure

Subjects were given written instructions describing the task that needed to be performed, and the attributes that need to be rated.

Our user study procedure was consistent with the ITU-R BT.2021 Recommendation, on subjective methods for the assessment of stereoscopic 3D systems [1]. For the experiment design, we have followed the double stimulus continuous quality scale (DSCQS) method. According to this procedure, subjects are shown a content, either test or reference; after a brief break, they are shown the other content. Then, both contents are shown for the second time, to obtain the subjective evaluations.This process is illustrated in Fig. 5.

To evaluate our method vis-à-vis the two other methods (DRC and Naive), we performed the tests in pairs of sessions

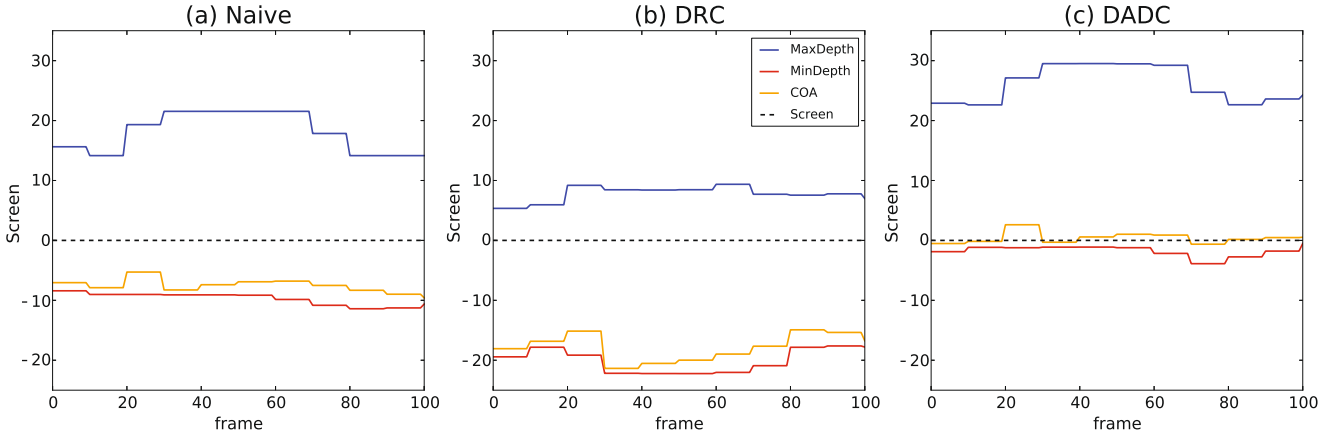


Fig. 6 Depth charts of an evaluated scene for the first hundred frames with (**a**) Naive method, (**b**) DRC, and (**c**) DADC

for each subject. For each pair of sessions, our method is used in the test content session while the compared method, either Naive or DRC, is used in the reference content session. The order of the reference and the test sessions in a pair and the order of the compared methods in consecutive pairs were both determined randomly. The subjects were not informed about either order. This set of tests were executed for each of our interactive scenes. Between the two sets of tests, a two minute break was introduced to relax eye muscles. Overall, eight test sessions were evaluated by each subject.

5.5 Assessment of contents

Subjects evaluated both test and reference content sessions of all cases separately, with respect to three criteria: quality, depth, and comfort. These three criteria are commonly used in the perceptual evaluation of stereoscopic contents [1]. The meaning of each criterion was explained to the subjects before the experiments. The motivation behind selecting these grading criteria is as follows:

- *Image Quality*: Image quality denotes the perceived overall visual quality of the shown content. Ghosting, defined as the incomplete fusion of the left and right image so that the image looks like a double exposure, is a critical factor determining the image quality of a stereoscopic content. A good quality 3D stereo image should eliminate the ghosting effect.
- *Perceived Depth*: This criterion measures the apparent depth as reported by the user, so that the effect of the methods on apparent depth should be taken into account.
- *Visual (Dis)comfort*: refers to the subjective sensation of discomfort that can be associated with improperly set stereoscopic parameters by the different algorithms. A good quality 3D stereo image should provide a comfortable viewing experience.

For assessment of the content, we also followed a methodology following the ITU-R BT.2021 Recommendation. We first asked the subjects to rate the quality, depth, and comfort of both the reference and test sessions separately, by filling out a 5-point Likert scale for each session. For assessment of quality, depth, and comfort, we used the discrete scale with the labels “bad”, “poor”, “fair”, “good”, and “excellent”. Then, at the end of each session pair, we also asked the subjects to compare between the two sessions. For this purpose, we asked the following questions in the evaluation form:

- Which session provided better image quality?
- Which session offered more depth?
- Which session was more comfortable to watch?
- Which session provided better overall quality?

5.6 Results

In order to analyze the user assessments, we computed the average scores for user ratings, as well as user preferences. Figure 7 illustrates the rating results for image quality, depth and comfort measures. The results show that our method yields better average than other approaches in all measures. Our DADC method achieved a considerable improvement particularly in the stereoscopic image quality, due to the fact that our method ensures the elimination of ghosting effect of the elements of interest in the scene to a significant extent. Regarding the assessment of image depth, the average rating of our method is slightly better than the other two methods, but less number of subjects have evaluated the depth impression of our method as “bad” or “poor”, compared to the other methods. The comfort ratings also reveal that our method is generally rated better than the other methods.

Figure 8 shows results of the preferences collected from the questions comparing our method with other methods described in Sect. 4. Different from the rating analysis of the

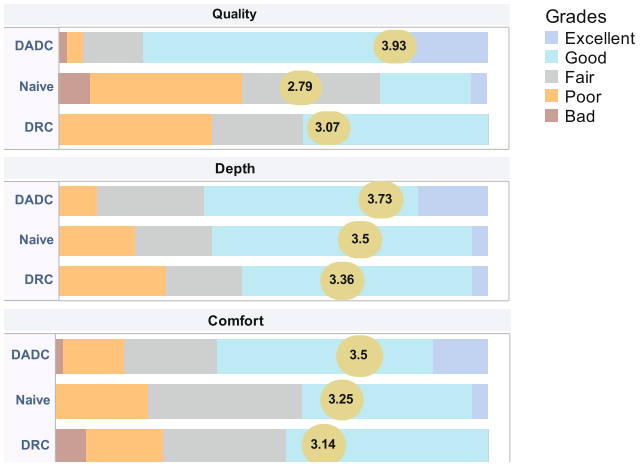


Fig. 7 Charts describing the subjects' ratings and averages based on 5-point Likert scale for our method and the compared methods. In each chart, the average grade is indicated in a circle

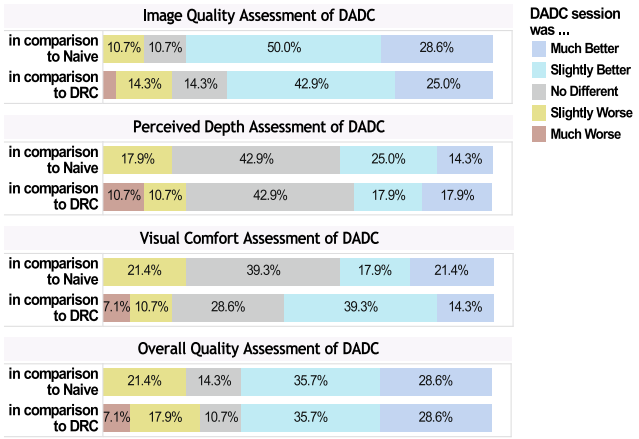


Fig. 8 Aggregated results from our session comparison questionnaires demonstrating relative user preferences of our DADC method in percentages. Scores are relative to Naive method in the first row and DRC method in the second

methods, this chart shows the preferences in percentages for our method directly in comparison with other two methods. These preferences are determined by the subjects by taking into account image quality, 3D perceived depth, visual comfort and overall quality. The study showed that DADC was preferred in overall quality over the two other methods, both with a 64.28 % preference; whereas in 21.43 % of the cases the Naive method was preferred over ours and 25 % showed preferences of DRC. The high performance of the Naive method is due to the fact that the static disparity levels were initialized compatibly with the scenes, for a fair comparison.

To evaluate the cinematographic quality of each method, we have plotted *the depth charts* [13] of a test sequence illustrating the distribution of the depth budget over time with each method. The charts in Fig. 6 shows the minimum and maximum depth values of the scene, with respect to the physical display surface (Fig. 2). The figure also shows the perceived depth of the most salient scene element, which we designated based on the scene and the significance scores (orange curve). The results show that our method achieves the goal of keeping the most significant object closed to the planar screen as much as possible. Based on these results, we can claim that our method prevents the accommodation/convergence conflict to a large extent.

6 Conclusion

This paper has presented a new approach for conveying scene depth in any arbitrary interactive 3D scene content by automatically calculating the stereoscopic camera parameters of convergence and camera separation. Our method specifies a depth configured according to the distribution and importance degree of salient elements in the scene, and automatically finds the parameters for mapping total scene depth to this specified depth range.

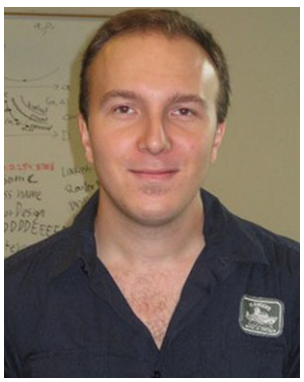
This new method for stereoscopic camera parameter arrangement allows 3D scene content creators to adjust and distribute available perceived depth in a way that the perceived depth is controlled and limited to the stereoscopic comfort zone of the users and accommodation/convergence conflict is not violated by keeping the focus or the convergence of the camera closer to the elements of interest.

Acknowledgements We would like to thank Dr. Ugur Gudukbay, Aytek Aman and Ates Akaydin for supplying some of the 3D human models used in our scenes; Sami Arpa and the 3dios Productions for providing the 3D display equipment; and also the anonymous reviewers for their valuable suggestions. This work is supported by the Scientific and Technical Research Council of Turkey (TUBITAK, project number 110E029).

References

1. Recommendation itu-r bt. 2021: Subjective methods for the assessment of stereoscopic 3dtv systems (2012)
2. Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H.P.: A perceptual model for disparity. ACM Trans. Graph. **30**(4) (2011). doi:10.1145/2010324.1964991 (Proceedings SIGGRAPH 2011, Vancouver)
3. Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H.P., Matusik, W.: A luminance-contrast-aware disparity model and applications. ACM Trans. Graph. **31**(6), 184:1–184:10 (2012)
4. Greß, A., Guthe, M., Klein, R.: Gpu-based collision detection for deformable parameterized surfaces. Comput. Graph. Forum **25**, 497–506 (2006)
5. Guttmann, M., Wolf, L., Cohen-Or, D.: Semi-automatic stereo extraction from video footage. In: IEEE 12th International Conference on Computer Vision, pp. 136–142. IEEE Press, New York (2009)
6. Heinzle, S., Greisen, P., Gallup, D., Chen, C., Saner, D., Smolic, A., Burg, A., Matusik, W., Gross, M.: Computational stereo camera system with programmable control loop. ACM Trans. Graph. **30**, 94:1–94:10 (2011)

7. Howard, I.P., Rogers, B.J.: Seeing in Depth. Depth Perception, vol. 2. I Porteous, Toronto (2002)
8. Itti, L., Koch, C.: Computational modelling of visual attention. Nat. Rev. Neurosci. **2**(3), 194–203 (2001)
9. Johnson, S.: The nlopt nonlinear-optimization package (2011). http://ab-initio.mit.edu/nlopt
10. Jones, G.R., Lee, D., Holliman, N.S., Ezra, D.: Controlling Perceived Depth in Stereoscopic Images. SPIE Press, Bellingham (2001)
11. Koppal, S., Zitnick, C., Cohen, M., Kang, S.B., Ressler, B., Colburn, A.: A viewer-centric editor for 3d movies. IEEE Comput. Graph. Appl. **31**(1), 20–35 (2011)
12. Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3d. ACM Trans. Graph. **29**(4), 75:1–75:10 (2010)
13. Liu, C.W., Huang, T.H., Chang, M.H., Lee, K.Y., Liang, C.K., Chuang, Y.Y.: 3d cinematography principles and their applications to stereoscopic media processing. In: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, pp. 253–262. ACM Press, New York (2011)
14. Mendiburu, B.: 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Focal Press, Waltham (2009)
15. Milgram, P., Krüger, M.: Adaptation effects in stereo due to online changes in camera configuration. In: Proc. SPIE, Stereoscopic Displays and Applications III, vol. 1669-13 (1992). SPIE Press, Bellingham
16. Oskam, T., Hornung, A., Bowles, H., Mitchell, K., Gross, M.: Oscam-optimized stereoscopic camera control for interactive 3d. In: SA'11 Proceedings of the 2011 SIGGRAPH Asia Conference, vol. 30, p. 189. ACM Press, New York (2011)
17. Reddy, M.: Perceptually optimized 3d graphics. IEEE Comput. Graph. Appl. **21**(5), 68–75 (2001)
18. Runarsson, T.P., Yao, X.: Stochastic ranking for constrained evolutionary optimization. IEEE Trans. Evol. Comput. **4**(3), 284–294 (2000)
19. Shamir, A., Sorkine, O.: Visual media retargeting. In: ACM SIGGRAPH ASIA 2009 Courses, SIGGRAPH ASIA '09, pp. 11:1–11:13. ACM Press, New York (2009)
20. Zilly, F., Kluger, J., Kauff, P.: Production rules for stereo acquisition. Proc. IEEE **99**(4), 590–606 (2011)
21. Zilly, F., Muller, M., Kauff, P., Schafer, R.: Stan—an assistance system for 3d productions: from bad stereo to good stereo. In: 2011 14th ITG Conference on Electronic Media Technology (CEMT), pp. 1–6. IEEE Press, New York (2011)



Ufuk Celikcan received his B.S. degrees in Electrical Engineering and Physics from Bogazici University, Turkey in 2006. He then received his M.S. degree in Electrical Engineering at University of California, Riverside, USA in 2010. He is currently a final year M.S. student in Computer Science at Bilkent University. His research interests include computer animation, human motion synthesis and analysis, stereoscopy, and joint source-channel video coding.

Gokcen Cimen received her B.S. in Computer Science from Izmir Institute of Technology (Turkey) in 2010 and she is currently a final year M.S. at Bilkent University (Turkey) for Computer Graphics and Animation. Her current research interests include data-driven character animation, computer graphics, and motion analysis and synthesis.

E. Bengu Kevinc received her B.S. degree in the Department of Computer Engineering from Atilim University, Ankara, Turkey in 2010. She is currently an M.S. Student in the Department of Computer Engineering at Bilkent University, Ankara, Turkey. Her research interests include computer graphics, stereoscopic 3D, and perception driven graphics applications.

Tolga Capin is an assistant professor at the Department of Computer Engineering at Bilkent University. He has received his Ph.D. at EPFL (Ecole Polytechnique Federale de Lausanne), Switzerland in 1998. He has more than 30 journal papers and book chapters, 50 conference papers, and a book. His research interests include networked virtual environments, mobile graphics, computer animation, and human-computer interaction.