ARTICLE IN PRESS
JID: CAG [m5G;December 5, 2017;12:37]

Computers & Graphics 000 (2017) 1–9

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Graphics Interface 2017

Depth annotations: Designing depth of a single image for depth-based effects

Jingtang Liao*, Shuheng Shen, Elmar Eisemann

Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 1 September 2017
Revised 17 November 2017
Accepted 18 November 2017
Available online xxx

Keywords:
Depth map design
A single image
Depth-based effects

ABSTRACT

We present a novel pipeline to generate a depth map from a single image that can be used as input for a variety of artistic depth-based effects. In such a context, the depth maps do not have to be perfect but are rather designed with respect to a desired result. Consequently, our solution centers around user interaction and relies on a scribble-based depth editing. The annotations can be sparse, as the depth map is generated by a diffusion process, which is guided by image features. We support a variety of controls, such as a non-linear depth mapping, a steering mechanism for the diffusion (e.g., directionality, emphasis, or reduction of the influence of image cues), and besides absolute, we also support relative depth indications. In case that a depth estimate is available from an automatic solution, we illustrate how this information can be integrated in form of a depth palette, that allows the user to transfer depth values via a painting metaphor. We demonstrate a variety of artistic 3D results, including wiggle stereoscopy, artistic abstractions, haze, unsharp masking, and depth of field.

1. Introduction

Representing 3D content on a standard 2D display is difficult. This topic has been of much interest to artists, who learned over centuries how to use effective pictorial cues to enhance depth perception on a canvas. On a computer display, it is also possible to add animation for the purpose of an increased depth perception. The *Ken Burns effect* is a simple example that combines zooming and panning effects and is widely used in screen savers. For television and movie productions, this technique can be obtained by a rostrum camera to animate a still picture or object. In its modern variant, the foreground is often separated from the background, which requires a rudimentary segmentation. The resulting parallax effect leads to a strong depth cue, when the viewpoint is changing (Fig. 1). Today, with the help of image-manipulation software, such effects can be easily produced. However, the picture elements are only translated, which is very restrictive and leads to a reduced effectiveness.

When several views are available, image-based view interpolation [1] is more general. The perceived motion of the objects helps in estimating spatial relationships. Nonetheless, these techniques often require a special acquisition setup or a carefully produced input. Wiggle stereoscopy can be seen as a particular case of view interpolation, which simply loops left and right images of a stereo pair and can result in a striking parallax perception despite its simplicity (Fig. 2). These techniques all avoid special equipment, e.g., 3D glasses, and they even work for people with limited or no vision in one eye.

Alternatively, it is possible to use a single input image and warp it based on a depth map to produce stereo pairs. Yet, computing depth maps for a monocular image is an ill-posed problem. While important advances have been made [2–5], the methods are not failsafe. Furthermore, many depth-based effects require the possibility for manual adjustments, such as remapping the disparity range of stereoscopic images and video in production, live broadcast, and consumption of 3D content [6], or to modify a depth-of-field effect in an artistic manner [7], which is why we focus on a semi-automatic solution. We will show that a depth estimate, if available, can be beneficial as a starting point for our interactive depth-map design.

In this paper, we propose a new framework to generate a depth map for a single input image with the goal of supporting artistic depth-based effects to illustrate the spatial information in the image. We build upon the insight that a depth map does not have to be perfect for such applications but should be easily adjustable by a user, as this option allows fine-tuning of the artistic effect. Our results are illustrated with a variety of examples, ranging from depth-of-field focus control to wiggle stereoscopy. Additionally, with such a depth map at hand, it is

* Corresponding author.
E-mail address: j.liao@tudelft.nl (J. Liao).

https://doi.org/10.1016/j.cag.2017.11.005
0097-8493/

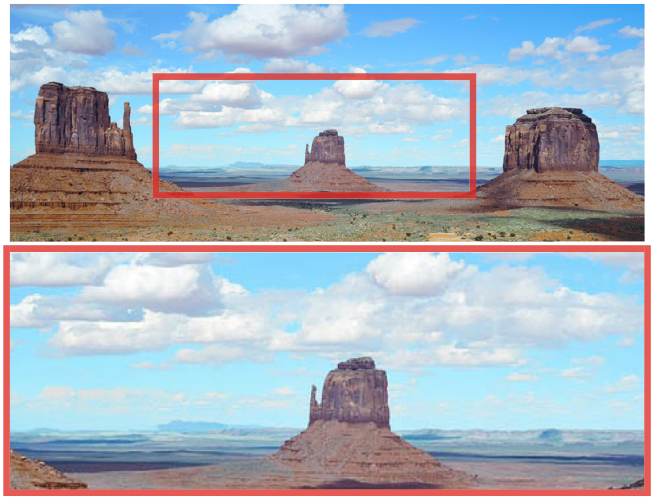

Fig. 1. Ken Burns effect. Panning and zooming on still images. (Image source: http://maxpixel.freegreatpicture.com.).

Fig. 2. Wiggle stereoscopy. Looping a left/right image pair (Image source: Wikimedia Commons). .

possible to produce image pairs for 3D viewing without (e.g., via establishing a cross-eyed view) or with specialized equipment (e.g., stereo glasses).

Our approach builds upon the assumption that depth varies mostly smoothly over surfaces and only exhibits discontinuities where image gradients also tend to be large. In consequence, we follow previous work and require only coarse annotations, such as sparse scribbles [8–10] or points [11]. These annotations form hard constraints in an optimization system that leads to a diffusion process, taking the image content into account. We focus on the control of this process and our method offers ways to influence the result via local and global constraints, such as defining relative depth differences, a non-linear depth diffusion by assigning a strength to scribbles, or privileged diffusion directions. We ensure that all these elements can be formulated in a linear optimization problem to ensure a fast solving step. We additionally show a selection of effects in our results.

This article presents an improved and extended version of [12]. Besides all relevant aspects of previous work, we introduce several new contributions. The original contributions include:

- A fast depth-map creation solution from a single image;
- Various tools to refine the depth map;
- A selection of effective effects, including wiggle stereography.

In this work, we also present the following extensions:

- A new depth design tool, in form of a depth palette if an estimated depth map is available.
- Additional depth-based effects, such as unsharp masking, haze, or new artistic abstractions.
- An extended discussion of the presented techniques and new experiments.

Furthermore, we describe new interface decisions to ease the creation of the depth map and facilitate the choice of adequate depth values.

2. Related work

Depth perception helps us perceive the world in 3D using various depth cues, classified into binocular and monocular cues. In an image, we typically encounter monocular cues — depth information that can be perceived with just one eye. Motion parallax [13], size, texture gradient [14], contrast, perspective, occlusion [15], and shadows [16] are examples of these. Motion parallax and occlusion are particularly strong [17]. Parallax arises due to the non-linear displacement relative to the depth when shifting the viewpoint of a perspective projection. In order to add such an effect, one can warp an image based on a depth map, which associates to each pixel the distance to the camera.

Depth estimation for a single image is a well-known problem in computer graphics and computer vision that received much attention. Recent approaches [2–5,18,19] are based on learning techniques. They enable an automatic conversion from a photo to a depth map. Nonetheless, the quality depends on the variety of the training data set and provided ground-truth exemplars. Additionally, in practice some manual segmentation is needed and the methods are not failsafe, as problematic elements are quite common (e.g., the reflections in a mirror or a flat image hanging on the wall). Even if an accurate depth is obtainable, it is not always optimal for artistic purposes [6,20], which is our focus.

Depth from defocus (DFD) is another approach where the amount of blur in different areas of a captured image is utilized to estimate the depth [21]. Methods for single DFD from conventional aperture are usually based on such assumptions. [22] assumed defocus blur to be the convolution of a sharp image with a 2D Gaussian function whose spread parameter is related to the object depth. [23] designed aperture filters based on texture sharpness. [24] took smoothness and color edge information into consideration to generate a coherent blur map for each pixel. [25] inferred depth information from photos by proposing a non-parametric matching prior with their constructed edgelet dataset, based on small small-scale defocus blur inherent in an optical lens. Their method is limited to photos in their original resolution and does not resolve ambiguities due to smooth edges. A general disadvantage of single-image DFD methods is that they cannot distinguish between defocus in front and behind the focal plane. Coded-aperture setups [26] address this issue by using a specially-designed aperture filter in the camera. [27] proposed an asymmetric aperture, which results in unique blurs for all distances from the camera. All these latter coded methods require camera modifications and have limitations regarding precision and image quality.

In our approach, the depth map will be designed by the user in a semi-automatic way. Hereby, also artistic modifications are kept possible. Early interactive techniques [28,29], and their extensions [30], focused on scenes containing objects with straight edges to reconstruct a 3D model by geometric reasoning and finding the best fitting model to line segments. In general, the use of edges is a good choice, as many natural scenes consist of smooth patches separated by object boundaries. [8] introduced a stroke-based user iterative framework in which users can draw a few sparse strokes to indicate depths as well as normals. Their technique optimizes for a smooth depth map in an edge-aware fashion, which is typically applied to photographs containing large planar geometry. [9] focused mainly on recovering depth maps for 2D paintings, where the 2D paintings have to be segmented into

Please cite this article as: J. Liao et al., Depth annotations: Designing depth of a single image for depth-based effects, Computers & Graphics (2017), https://doi.org/10.1016/j.cag.2017.11.005