Computers & Graphics 35 (2011) 99–111

Contents lists available at ScienceDirect

Computers & Graphics

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

Extended papers from NPAR 2010

Non-photorealistic, depth-based image editing

Jorge Lopez-Moreno [a,*], Jorge Jimenez [a], Sunil Hadap [b], Ken Anjyo [c], Erik Reinhard [d], Diego Gutierrez [a]

[a] *Universidad de Zaragoza, Spain*
[b] *Adobe Systems, Inc., USA*
[c] *OLM Digital, Inc., Japan*
[d] *University of Bristol, UK*

ARTICLE INFO

Article history:
Received 30 August 2010
Received in revised form
9 November 2010
Accepted 11 November 2010
Available online 19 November 2010

Keywords:
Non-photorealistic rendering
Relighting
Image processing
Human visual system
Real-time

ABSTRACT

Recent works in image editing are opening up new possibilities to manipulate and enhance input images. Within this context, we leverage well-known characteristics of human perception along with a simple depth approximation algorithm to generate non-photorealistic renditions that would be difficult to achieve with existing methods. Once a perceptually plausible depth map is obtained from the input image, we show how simple algorithms yield powerful new depictions of such an image. Additionally, we show how artistic manipulation of depth maps can be used to create novel non-photorealistic versions, for which we provide the user with an intuitive interface. Our real-time implementation on graphics hardware allows the user to efficiently explore artistic possibilities for each image. We show results produced with six different styles proving the versatility of our approach, and validate our assumptions and simplifications by means of a user study.

1. Introduction

Whether the goal is to convey a specific mood, to highlight certain features or simply to explore artistic approaches, non-photorealistic rendering (NPR) provides an interesting and useful set of techniques to produce computer-assisted stylizations. Most of those techniques either leverage 3D information from a model, work entirely in 2D image space, or use a mixed approach (typically by means of a Z- or G-buffer) [1]. We are interested in exploring new possibilities for stylized depiction using a single image as input, while escaping traditional limitations of a purely 2D approach. For instance, the design of lighting schemes is crucial to communicate a scene's mood or emotion, for which depth information is required.

Our key observation is the fact that a single photograph or painting has richer information than we might expect. In particular, we ask ourselves what layers of information present in an image may have been usually overlooked by stylized depiction techniques? And what would the simplest way to access that "hidden" information be, in a way that allows dramatic manipulation of the look of an image?

This paper presents a set of stylization techniques that deals with a single photograph as input. It is well known that when it comes to stylized depiction, human perception is able to build complex shapes with very limited information, effectively filling in missing detail whenever necessary, as illustrated in Fig. 1 (left). The power of suggestion and the influence of light and shadows in controlling the emotional expressiveness of a scene have also been extensively studied in photography and cinematography [2,3]: for instance, carefully placed shadows can turn a bright and cheerful scene into something dark and mysterious, as in Fig. 1 (right).

With this in mind, we propose a new class of methods for stylized depiction of images based on approximating significant depth information at local and global levels. We aim to keep the original objects recognizable while conveying a new mood to the scene. While the correct recovery of depth would be desirable, this is still an unsolved problem. Instead, we show that a simple methodology suffices to stylize 3D features of an image, showing a variety of 3D lighting and shading possibilities beyond traditional 2D methods, without the need for explicit 3D information as input. An additional advantage of our approach is that it can be mapped onto the GPU, thus allowing for real-time interaction.

Within this context, we show stylized depictions ranging from simulating the *chiaroscuro* technique of the old masters like Caravaggio [4] to techniques similar to those used in comics. In recent years, both the movie industry (Sin City, A Scanner Darkly, Renaissance, etc.) and the photography community (more than 4000 groups related to comic art on Flickr) have explored this medium. The goal of obtaining comic-like versions of photographs has even motivated the creation of applications such as Comic Life.[1]

* Corresponding author.
E-mail address: jlm@unizar.es (J. Lopez-Moreno).

[1] http://plasq.com/comiclife-win.

0097-8493/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2010.11.007

Fig. 1. Left: The classic image of "The Dog Picture", well known in vision research as an example of emergence: even in the absence of complete information, the shape of a dog is clearly visible to most observers (original image attributed to R. C. James [5]). Right: Example of dramatically altering the mood of an image just by adding shadows.

2. Previous work

Our work deals with artistic, stylized depictions of images, and thus falls under the NPR category. This field has produced techniques to simulate artistic media, create meaningful abstractions or simply to allow the user to create novel imagery [6,7]. In essence, the goal of several schools of artistic abstraction is to achieve a depiction of a realistic image, where the object is still recognizable but where the artist departs from the accurate representation of reality. In this departure, the object of depiction usually changes: a certain mood is added or emphasized, unnecessary information is removed and often a particular visual language is used.

In this paper, we explore what new possibilities can be made available by adding knowledge about how the human visual system (HVS) interprets visual information. It is therefore similar in spirit to the work of DeCarlo and Santella [8] and Gooch et al. [9]. DeCarlo and Santella propose a stylization system driven by both eye-tracking data and a model of human perception, which guide the final stylized abstraction of the image. Their model of visual perception correlates how interesting an area in the image appears to be with fixation duration, and predicts detail visibility within fixations based on contrast, spatial frequency and angular distance from the center of the field of view. Although it requires the (probably cumbersome) use of an eye-tracker, as well as specific per-user analysis of each image to be processed, the work nevertheless shows the validity of combining perception with NPR techniques, producing excellent results.

Instead, we apply well-established, general rules of visual perception to our model, thus freeing the process from the use of external hardware and individual image analysis. The goals of both works also differ from ours: whilst DeCarlo and Santella aim at providing meaningful abstraction of the input images, we are predominantly interested in investigating artistic possibilities.

Gooch and colleagues [9] multiply a layer of thresholded image luminances with a layer obtained from a model of brightness perception. The system shows excellent results for facial illustrations. It is noted that in their approach some visual details may be difficult (or even impossible) to recover. Although in the context of facial stylization this counts as a benefit, it might not be desirable for more general imagery.

Depth information has previously been used to aid the generation of novel renditions. For instance, ink engravings can be simulated by estimating the 3D surface of an object in the image, and using that to guide strokes of ink [10]. This method is capable of producing high-quality results, although it requires the user to individually deform 3D patches, leading to a considerable amount of interaction. The algorithms proposed by Oh et al. [11] cover a wide range of image scenarios with specific solutions to extract 3D data for each one, but also come at the expense of considerable manual input. Okabe and colleagues [12] present an interactive technique to estimate a normal map for relighting, whereas in [13], painterly art maps (PAMs) are generated for NPR purposes. While both works show impressive results, they again require intensive, skilled user input, a restriction we lift in our system.

In their work, Raskar and colleagues [14] convey shape features of objects by taking a series of photographs with a multi-flash camera strategically placed to cast shadows at depth discontinuities. Akers et al. [15] take advantage of relighting to highlight shape and features by combining several images with spatially varying light mattes, while in [16] details are enhanced in 3D models via exaggerated shading. In contrast, our approach operates on single off-the-shelf images, allows for new, artistic lighting schemes, and requires at most a user-defined mask to segment objects, for which several sophisticated tools exist [17,18].

In the field of halftone stylization based on 3D geometry we should mention the recent work of Buchholz et al. [19], which incorporates information from shading, depth and geometry in order to generate boundaries between black and white regions which run along important geometric features for shape perception (like creases).

Bhat et al. [20] proved the potential of gradient-based filtering in the design of image processing algorithms like painterly rendering or subtle image relighting.

A 2.5D approach has been explored in the context of video stylization [21], aiding the production of hatching and painterly effects. This method, however, requires the specific calibrated capture of the 2.5D video material to be processed, which is still either cumbersome or expensive.

Lopez-Moreno et al. [22] showed that 2.5D approximations suitable for NPR can be obtained from off-the-shelf images by applying current theories about the perception of shape. In the present paper we have extended their work in a two-fold manner: First, we have explored two new stylization methods based on more complex rendering methods; ambient occlusion and global illumination. And second, we have developed an interactive editing interface which complements the edition of lighting by providing the user with full artistic control over the generation of color, shading and shadows.

3. Perceptual background

At the heart of our algorithm, which will be described in the next section, lies the extraction of *approximate* depth information from the input image. Since we do not have any additional information other than pixel values, we obviously cannot recover depth accurately, and therefore the result will potentially contain large errors. However, given that we are interested in stylized depictions of images, we will show that we do not require physical accuracy, but only approximate values which yield pleasing, plausible results. Our depth approximation algorithm leverages some well-known characteristics of the human visual system. Although the inner workings of human depth perception are not yet fully understood, there exist sufficient indicators of some of its idiosyncracies that enable us to approximate a reasonable depth map for our purposes. In particular, we rely on the following observations:

1. Belhumeur et al. [23] showed that for unknown Lambertian objects, our visual system is not sensitive to scale transformations along the view axis. This is known as the *bas-relief ambiguity*, and due to this implicit ambiguity large scale errors